Field Guide to

Lens Design

Julie Bentley
Craig Olson

SPIE Field Guides
Volume FG27

John E. Greivenkamp, Series Editor

Introduction to the Series

Welcome to the *SPIE Field Guides*—a series of publications written directly for the practicing engineer or scientist. Many textbooks and professional reference books cover optical principles and techniques in depth. The aim of the *SPIE Field Guides* is to distill this information, providing readers with a handy desk or briefcase reference that provides basic, essential information about optical principles, techniques, or phenomena, including definitions and descriptions, key equations, illustrations, application examples, design considerations, and additional resources. A significant effort will be made to provide a consistent notation and style between volumes in the series.

Each *SPIE Field Guide* addresses a major field of optical science and technology. The concept of these *Field Guides* is a format-intensive presentation based on figures and equations supplemented by concise explanations. In most cases, this modular approach places a single topic on a page, and provides full coverage of that topic on that page. Highlights, insights, and rules of thumb are displayed in sidebars to the main text. The appendices at the end of each *Field Guide* provide additional information such as related material outside the main scope of the volume, key mathematical relationships, and alternative methods. While complete in their coverage, the concise presentation may not be appropriate for those new to the field.

The *SPIE Field Guides* are intended to be living documents. The modular page-based presentation format allows them to be easily updated and expanded. We are interested in your suggestions for new *Field Guide* topics as well as what material should be added to an individual volume to make these *Field Guides* more useful to you. Please contact us at **fieldguides@SPIE.org**.

John E. Greivenkamp, *Series Editor*
Optical Sciences Center
The University of Arizona

Field Guide to Lens Design

Optical design has a long and storied history, from the magnifiers of antiquity, to the telescopes of Galileo and Newton at the onset of modern science, to the ubiquity of modern advanced optics. The process for designing lenses is often considered both an art and a science. While advancements in the field over the past two centuries have done much to transform it from the former category to the latter, much of the lens design process remains encapsulated in the experience and knowledge of industry veterans. This *Field Guide* provides a working reference for practicing physicists, engineers, and scientists for deciphering the nuances of basic lens design. Because the optical design process is historically (and quite practically) closely related to ray optics, this book is intended as a companion to the *Field Guide to Geometrical Optics*, in which first-order optics, thin lenses, and basic optical systems are treated in more detail. Note that this compact reference is not a substitute for a comprehensive technical library or the experience gained by sitting down and designing lenses.

This material was developed over the course of several years for undergraduate and graduate lens design classes taught at the University of Rochester. It begins with an outline of the general lens design process before delving into aberrations, basic lens design forms, and optimization. An entire section is devoted to techniques for improving lens performance. Sections on tolerancing, stray light, and optical systems are followed by an appendix covering related topics such as optical materials, nonimaging concepts, designing for sampled imaging, and ray tracing fundamentals, among others.

Thanks to both of our families—Danielle, Alison, Ben, Sarah, Julia, and especially our spouses, Jon and Kelly. The cats will now get fed, and all soccer parents beware!

<div align="right">

Julie Bentley
University of Rochester

Craig Olson
L-3 Communications

</div>

Table of Contents

Table of Contents

Table of Contents

Table of Contents

Glossary of Symbols and Acronyms

A	Area
AOI	Angle of incidence
AR	Antireflection
BBAR	Broadband antireflection coating
BFL	Back focal length
BFS	Best fit sphere
BRDF	Bidirectional reflectance distribution function
BSDF	Bidirectional scattering distribution function
BTDF	Bidirectional transmittance distribution function
c	Surface curvature
C	Lens conjugate factor
CA	Clear aperture
CCD	Charge-coupled device
CDF	Cumulative distribution function
CGH	Computer-generated hologram
CMOS	Complementary metal-oxide semiconductor
CRA	Chief ray angle
CT	Center thickness
CTE	Coefficient of thermal expansion
CTF	Contrast transfer function
d	Airspace
d	Thickness
DLS	Damped least squares
dn/dT	Thermo-optic coefficient
DOE	Diffractive optical element
EFL	Effective focal length
EPD	Entrance pupil diameter
ESF	Edge-spread function
ETD	Edge thickness difference
f	Focal length
$f/\#$	f-number or relative aperture
FEA	Finite-element analysis

Glossary of Symbols and Acronyms

FFL	Front focal length
FFOV	Full field of view
FFT	Fast Fourier transform
FOV	Field of view
GQ	Gaussian quadrature
GRIN	Gradient index
h, h'	Object/image height
H	Lagrange invariant
H	Normalized field coordinate
HFOV	Half field of view
HO	Higher order
HOE	Holographic optical element
HR	High-reflection
i, i'	Angle of incidence w.r.t. surface normal
i, i_a	Marginal ray angle w.r.t. surface normal
$\bar{i}, i_b$	Chief ray angle w.r.t. surface normal
ID	Inner diameter of a lens barrel or mount
IR	Infrared
L	Radiance
LOS	Line of sight
LR	Limiting resolution
LSF	Line spread function
LWIR	Long-wave infrared
m	Diffraction order
m	Magnification
MP	Magnifying power (magnifier or telescope)
MTF	Modulation transfer function
MWIR	Midwave infrared
n, n'	Index of refraction
$n(z), n(r)$	Gradient index profile function
NA	Numerical aperture
NITD	Narcissus-induced temperature difference
NRT	Nonsequential ray tracing
NUC	Nonuniformity correction

Glossary of Symbols and Acronyms

OAP	Off-axis parabola
OAR	Off-axis rejection
OD	Outer diameter (of a lens)
OPD	Optical path difference
OTF	Optical transfer function
p	Pixel pitch in sampled detector arrays
P	Partial dispersion
PDF	Probability distribution function
PSF	Point spread function
PSNIT	Point-source normalized irradiance transmittance
PST	Point-source transmittance
P–V	Peak to valley
Q	Sampling ratio
r	Radial surface coordinate
R, ROC	Radius of curvature
RI	Relative illumination
RMS	Root mean square
RSS	Root sum square
RT	Reflective triplet
s, s'	Object/image distance
SA	Spherical aberration
SLR	Single-lens reflex
t	Thickness or airspace
T	Temperature
TIR	Total indicator runout
TIR	Total internal reflection
TIS	Total integrated scatter
TMA	Three-mirror anastigmat
TML	Three mirror long
u, u'	Paraxial ray angles w.r.t. optical axis
u, u_a	Marginal ray angle w.r.t. optical axis
$\bar{u}$, u_b	Chief ray angle w.r.t. optical axis
UV	Ultraviolet
V	Abbe number
W	Wave aberration function

Glossary of Symbols and Acronyms

W_{ijk}	Wavefront aberration coefficient
WD	Working distance
y, y_a	Marginal ray height at a surface
$\bar{y}$, y_b	Chief ray height at a surface
z	Optical axis
$z(r)$	Surface sag/profile function
Z_n	Zernike polynomial coefficient
β	Lens shape factor
$\Delta\lambda$	Wavelength range or bandwidth
δz	Defocus
ε	Obscuration ratio
$\varepsilon, \varepsilon_x, \varepsilon_y$	Transverse ray error
θ, θ'	Angle of incidence/refraction
θ	Half field of view
θ	Pupil azimuthal coordinate
κ	Conic constant
λ	Wavelength
λ_0	Center wavelength
ρ, ρ_x, ρ_y	Normalized radial pupil coordinate
Φ	System power
ϕ	Element or surface power
ϕ	Merit or penalty function

Sign Conventions

Throughout this **Field Guide**, a set of fully consistent sign conventions is utilized.
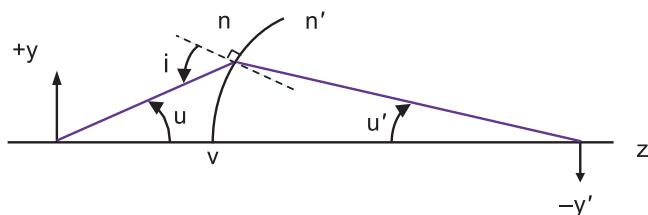
- The z axis is the **optical axis** or the axis of symmetry of a rotationally symmetric optical system.
- The y axis is in the plane of the paper, perpendicular to the z axis, and the x axis points directly into the paper, forming a right-handed coordinate system.
- All distances are measured relative to a reference point, line, or plane in a Cartesian sense: **directed distances** above or to the right are positive; below or to the left are negative.
- Angles u are measured from the optical axis to the ray with the smallest of the two angles chosen, while **angle of incidence** i are measured from the normal of the surface to the incident ray.
- Counter-clockwise angles are positive, and clockwise angles are negative.
- The **radius of curvature** of a surface is defined as the directed distance from its vertex to its center of curvature and is therefore positive when the center of curvature is to the right of the surface.
- Light travels from left to right (from $-z$ to $+z$) unless reflected.
- A $'$ (prime) symbol following a variable denotes a value after refraction at a surface.
- The signs of all indices of refraction following a reflection are reversed.
- To aid in the use of these conventions, all directed distances and angles are identified by arrows with the tail of the arrow at the reference point, line, or plane.

Basic Concepts

A **lens** is often described as a device that creates an image of an object on a detector. More generally, it can be thought of as any system that tries to collect and distribute light in a specified way. Although a lens can consist of a single lens element, it is often composed of multiple lens elements of different shapes and sizes, referred to as a **lens system** or an **optical system**.

Optical design or **lens design** is the process used to find the best set of lens construction parameters (e.g., radii of curvature, thicknesses, airspaces, and materials) that optimizes the overall performance (including the manufacturability) of an optical system.

A **specification document** lists the requirements needed to design a lens system. It defines basic specifications such as object and image location, aperture, field of view, and wavelength range. It contains image quality measures such as spot size, resolution, and distortion. It also includes packaging and environmental requirements such as diameter, length, and temperature.

Optical design software is computer code that aids in the selection and optimization of lens construction parameters and performance evaluation. Given a starting point and a set of variables and constraints, the computer program drives an optical design to a local optimum, as defined by a merit function.

A **lens designer** (or "ray bender") finds the variations in lens parameters that yield the greatest improvements in optical system performance. Computer optimization and real **ray tracing** (use of Snell's law to find the position and direction of a ray after refraction or reflection at each lens surface) are key components of the process. A lens designer is also often responsible for developing a specification document (and resolving any conflicts) and identifying engineering trades between performance, packaging constraints, tolerances, cost, and schedule.

Optical Design Process

The basic **optical design process** is as follows:

- A **specification document** for the optical system is established, and the appropriate type of lens form (e.g., microscope, telescope, camera lens) is selected.

- A starting-point design is chosen from previous projects, historical databases, or patent literature (primarily based on the aperture, field, wavelength, and packaging requirements).

- If an appropriate starting-point design does not exist, designers can choose to start from a **first-order solution** that identifies focal lengths and object/image/pupil locations and sizes using thin lenses and paraxial rays. The first-order solution can be converted to a third-order **thin lens** solution where thin lens parameters such as radius of curvature and index of refraction are chosen to minimize **third-order aberrations**. Thickness is then introduced into the thin lenses to arrive at a **thick lens** starting point.

- The starting-point design is assessed by tracing real rays at multiple wavelengths through the system to analyze the chromatic and higher-order aberrations. Compliance to performance requirements (e.g., image quality, packaging restrictions, fabrication and assembly tolerances, cost) is evaluated.

- If the performance is not acceptable, **optimization** is used to improve the design. A set of variables and constraints is defined, and computer software is used to find design solutions as defined by a **merit function**.

- If the solution is still unacceptable, designers must return to a previous step and iterate until the design meets specifications. This process may require the selection of a different starting-point design and/or requesting a significant change in one or more of the specifications.

- The manufacturability of the design is just as important as the final image quality. Therefore, a complete design process and performance evaluation includes **tolerancing**, **stray light analysis**, and **thermal** and other **environmental analyses**.

Aperture and Wavelength Specifications

The **aperture specification** determines the amount of light collected from an object and establishes a resolution limit to the system. The three most common ways to specify an aperture for a lens system are **entrance pupil diameter (EPD)**, *f*/#, and **numerical aperture (NA)**.
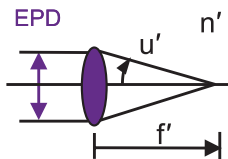
The entrance pupil diameter defines the size of the beam in object space and is a common way to specify the aperture of lens systems during computer-aided design. The ratio of the focal length of a lens to its entrance pupil diameter is called the **relative aperture** or *f*/#.

$$f/\# = \frac{f}{EPD} \qquad NA = n' \sin u'.$$

Numerical aperture is the index of refraction of the image space media times the sine of the half-angle cone of light. *f*/# is typically used to specify systems with distant objects (e.g., camera lenses or telescopes), whereas NA is used to specify systems that work at finite conjugates (e.g., relay systems or microscope objectives).

For infinite conjugate systems, *f*/# and NA are related by *f*/# = 1/(2*NA*). Although this equation only holds for systems with infinite object distances, the same quantity is often calculated for finite conjugate systems and is referred to as the **working *f*/#**.

The spectral range of an optical system is identified by a center wavelength λ_0 and a **bandwidth** $\Delta\lambda$. **Relative bandwidth** is the ratio of the total bandwidth to the center wavelength. Narrow-bandwidth systems are called **monochromatic** (e.g., laser collimators), whereas larger-bandwidth systems are called **polychromatic** (e.g., camera objectives). The optical materials are ideally chosen for both high transmission at the system wavelengths and good chromatic correction over the full bandwidth.

Resolution and Field of View

$$\Delta X = \frac{0.61\lambda}{NA}$$
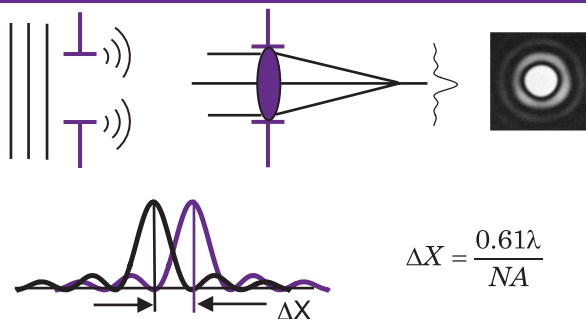
Diffraction from a finite circular aperture results in an **Airy disc** pattern. The **Rayleigh criterion** determines the smallest lateral separation ΔX between two Airy disc patterns for which two distinct points can still be resolved. Therefore, the fundamental **resolution** of a rotationally symmetric optical system is a function of both its wavelength and its numerical aperture. To increase resolution, systems move to shorter wavelengths and higher numerical apertures.

The **field of view (FOV)** of an optical system is specified as an object height, object angle, or image height. The image height can be directly correlated to the size of the detector (e.g., film or CCD) but can also represent the size of the field over which the lens is well corrected. The **half field of view (HFOV)** is the *radius* of the object height h or image height h', whereas the **full field of view (FFOV)** indicates a *diameter* measure.
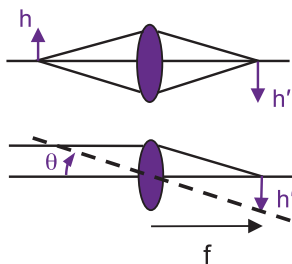
For lenses with infinite object distances, one can quickly convert from angular HFOV θ to image height h' using
$$h' = f \tan\theta$$
where f is the focal length of the system. This equation is also very useful for converting detector resolution $\Delta h'$ to angular resolution $\Delta\theta$ for a focal length f.

Packaging and Environment

The achievable imaging performance of a lens design is often limited by **packaging** and **interface requirements**. For example, the maximum number of elements that can be used in a design may be dictated by a maximum lens **envelope** (diameter and length) and/or weight constraint. Other interface requirements that may directly or indirectly drive the design include working distance, pupil location and size, filter/window/fold mirror position and size, and center-of-gravity location and shift (for lenses with movable elements). Special access for modularity (e.g., an eyepiece) or field service may also be required.
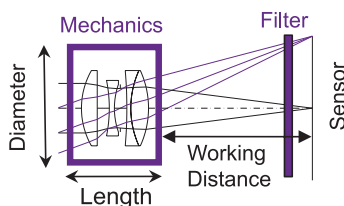
Most lens systems have a **working distance** requirement to keep the lens from crashing into an object or a sensor. It is important to clarify if a specification references the vertex of the lens element (**back focal length** or **BFL**), the closest mechanical surface point to the sensor (**image clearance**), or a mechanical flange surface (**flange-to-focus**).

Environmental conditions and outside influences such as **temperature**, **altitude**, and **vibration** place additional constraints on a lens design. **Moisture** or **humidity** effects may induce coating delamination, absorption and scattering from condensed water droplets, and glass staining. **Shock requirements** dictate single-event extreme conditions that the lens must survive (e.g., a thermal shock, drop, or sudden pressure loss that can occur during shipping). Extreme environments, such as highly corrosive salt water, or use in a vacuum chamber can further restrict design freedom. Lenses operating in ionizing radiation require special **radiation-hardened** glass choices because many optical glasses will "fog" or become opaque over time in such an environment.

Wave Aberration Function

Geometrical optics describes the properties of lenses when the effects of diffraction and interference are ignored (in the limit where the wavelength approaches zero). In this case, an object consists of a collection of independently radiating point sources. Each point source emits a spherical **wavefront** (surface of constant optical path length from a point on the object), and rays are traced perpendicular to the wavefront.

First-order optics is the study of perfect optical systems without aberrations. In a perfect system, an incident spherical wave is mapped to a perfect spherical wave in the exit pupil, converging on the ideal image location.
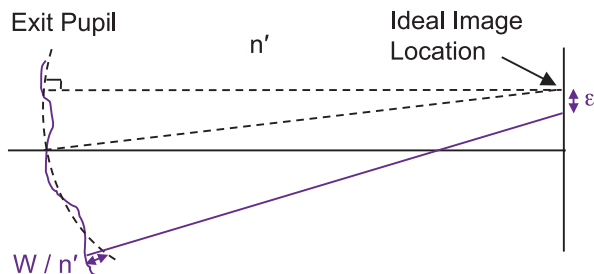
Aberrations are imperfections in the image formation of an optical system. They are measured either by the deformation of the wavefront in the exit pupil or by the lateral distance ε, by which the rays miss the ideal image point. The distance ε is called the **transverse ray error**, whereas the **wave aberration function** W is the optical length, measured along a ray, from the aberrated wavefront to the reference sphere.

W and ε are related by a partial derivative with respect to pupil coordinate ρ, where u'_a is the angle of the marginal ray in image space:

$$\varepsilon'_y = \frac{1}{n'u'_a}\frac{\partial W}{\partial \rho_y} \qquad \varepsilon'_x = \frac{1}{n'u'_a}\frac{\partial W}{\partial \rho_x}$$

Third-Order Aberration Theory

In a rotationally symmetric optical system, one can specify any ray in the system with only three coordinates: its 1D position in the field plane (typically along the *y* axis), denoted by the **normalized field coordinate** *H*, and its 2D position in the pupil plane, denoted by the **normalized pupil coordinates** $\rho_x(\rho\sin\theta)$ and $\rho_y(\rho\cos\theta)$.
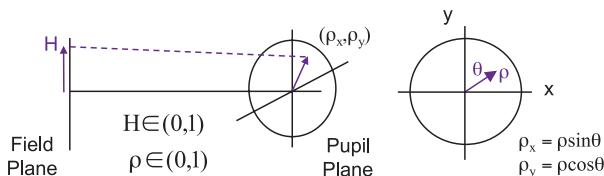
Monochromatic aberrations are described by expanding the wave aberration function *W* in a power series of aperture and field coordinates, ρ, θ, and *H*:

$$W_{IJK} \Rightarrow H^I \rho^J \cos^K \theta$$

$$W(H,\rho,\theta) = W_{020}\rho^2 + W_{111}H\rho\cos\theta + W_{040}\rho^4 + W_{131}H\rho^3\cos\theta$$
$$+ W_{222}H^2\rho^2\cos^2\theta + W_{220}H^2\rho^2 + W_{311}H^3\rho\cos\theta + O(6)$$

where the other terms in the power series are forbidden by rotational symmetry. In **third-order aberration theory** (because of the derivative dependence, *third*-order transverse aberrations have a *fourth*-order wavefront dependence), the higher-order terms *O*(6) in the wave expansion are dropped. The coefficients for the remaining terms are traditionally defined as

W_{020}: Defocus W_{222}: Astigmatism
W_{111}: Wavefront tilt W_{220}: Field curvature
W_{040}: Spherical aberration W_{311}: Distortion
W_{131}: Coma

The value of each polynomial coefficient (W_{IJK}) gives the magnitude of the aberration and is a function of lens parameters (e.g., index of refraction, thickness, and radius of curvature), and marginal/chief ray heights and angles.

Spot Diagram and Encircled Energy

A **spot diagram** or **scatter plot** shows the transverse ray error (ε_y versus ε_x) for a given object height H. Rays are traced into the entrance pupil from a single object point. A "spot" is plotted where each ray hits the image plane with respect to the paraxial image point; several hundred rays at different wavelengths are usually plotted.
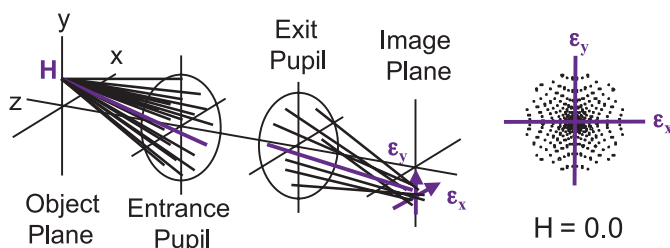
Spot diagrams are good for final system evaluation, especially for systems with aberrations much larger than the diffraction limit.

The **RMS spot diameter** or **RMS spot size** is a common metric for the quality of an image point. Many rays are traced from a single object point, and the **centroid** or "center of gravity" of all of the image plane intersections is determined. The RMS spot size is then calculated by taking the square root of the sum of the squares of each ray's distance from the centroid, divided by the total number of rays.

Encircled energy is an image quality measure for a single field point, using the intensity distribution of a spot to calculate the percentage of transmitted energy through an opening of variable radius. **Ensquared energy** assumes that the opening has a square shape.
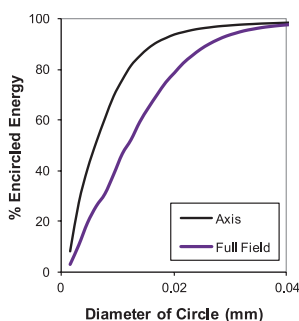
Transverse Ray Plot

Transverse ray aberration plots or **ray intercept plots** are plots of transverse ray error versus the relative position of the ray in the pupil (ε_y versus ρ_y, and ε_x versus ρ_x) for different field points H. They are similar to spot diagrams, but instead of tracing a 2D grid of rays in the pupil, they use a **tangential** (Y) and a **sagittal** (X) slice of the pupil. A "perfect" lens has a straight line along the horizontal axis in both plots.

● Tangential Rays
▲ Sagittal Rays
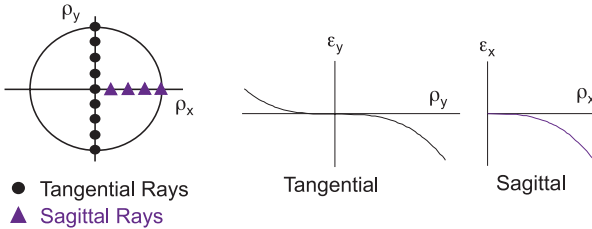
Transverse ray plots and spot diagrams are typically plotted for three different field points: on-axis, 0.7 field, and full-field. The shape of the plots at different values of H yield a lot of information about which aberrations are present in the lens. An estimate of peak-to-valley spot diameter (Δ) can be easily calculated from the transverse ray plot by drawing a rectangle around the total error.
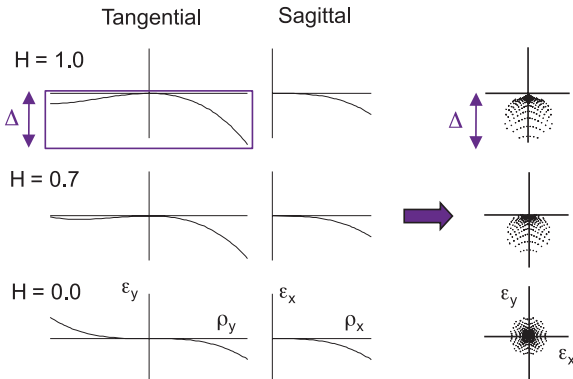
Wavefront or OPD Plots

The departure of an aberrated wavefront from an ideal spherical wavefront is called the **optical path difference (OPD)** and is typically measured in units of wavelength. **Wavefront** or **OPD plots** are plots of OPD versus ρ_y and OPD versus ρ_x for different field points H. Wavefront plots are proportional to the integral of the transverse ray error plot and also indicate what aberra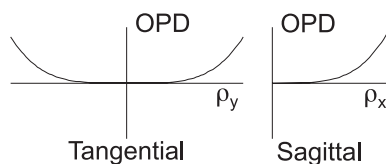tions are present. OPD plots are used instead of transverse ray plots for well-corrected systems with nearly diffraction-limited performance.

> Wavefront plots show slices of the full wavefront error of an optical system that would be measured by a test interferometer at a given field point.

Peak-to-valley (P–V) OPD is the difference between the wavefront maximum and the wavefront minimum, and is a good measure of the im-

age quality of an optical system if the wavefront error is smooth with small changes in slope. Most real wavefronts are not smooth, and the **RMS OPD** is generally a better measure of image quality. The RMS OPD is calculated by taking the square root of the mean of the squared OPD values sampled over the pupil.

The **Rayleigh quarter-wave criterion** states that if the total wavefront deviation is less than a quarter of a wavelength (P–V OPD $< \lambda/4$), then the image quality is "good" or nearly diffraction limited. Although the exact relationship between P–V OPD and RMS varies with the shape of the wavefront, a common rule of thumb is RMS OPD = (P–V OPD/5). This is then used to define the RMS limit, RMS OPD $< \lambda/20$ (0.05), for the image quality of a "good" lens.

Point Spread Function and Strehl Ratio

The **point spread function (PSF)** is the impulse response function of a lens and includes the effects of both aberrations and diffraction. For a single point object, a PSF plot shows the full 3D intensity distribution in the image plane. In practice, a **line spread function (LSF)** or **edge spread function (ESF)** gives the intensity distribution of a line or edge object and is a more-convenient measure of the lens performance.
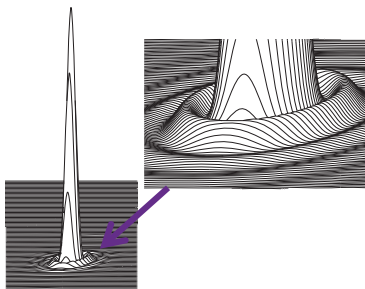
The **Strehl ratio** is defined as the ratio of the peak intensity of an aberrated lens to the peak intensity of an aberration-free lens. The Strehl ratio is a good single-number, image-quality design measure for well-corrected systems but is less meaningful for highly aberrated systems.
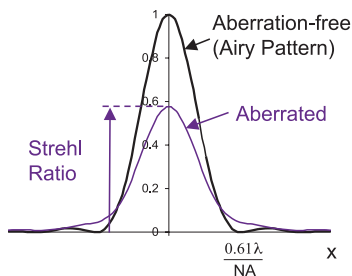
The **Maréchal criterion** for image quality is a Strehl ratio of 0.82, which is equivalent to an RMS OPD of $\sim \lambda/14$ or 0.071λ RMS. A Strehl ratio of 0.8 corresponds closely to a P–V OPD of $\lambda/4$ for a lens with only defocus and/or spherical aberration. Both criteria are commonly referred to as **diffraction limited**.

For a lens with a relatively small OPD, the relationship between the Strehl ratio and the RMS OPD is approximated by

$$Strehl \cong \left(1 - 2\pi^2 \omega^2\right) \qquad \omega = RMS_{OPD}$$

MTF Basics

$$\text{Contrast} = \frac{I_{max} - I_{min}}{I_{max} + I_{min}}$$

Phase Reversal
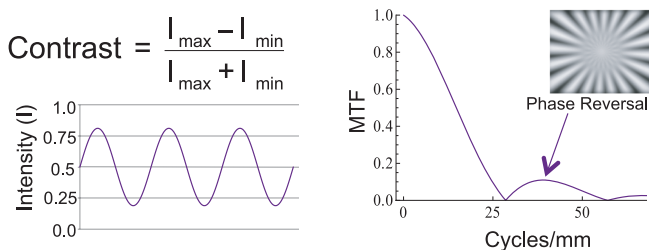
The **modulation transfer function (MTF)** is the magnitude of the complex **optical transfer function (OTF)** and is a measure of how well a lens relays contrast from object to image. MTF is the ratio of the image sine wave modulation to that of the object as a function of the **spatial frequency**, typically specified at the image in **cycles/mm**. A cycle is one full period (peak and trough) of the sine wave. For distant objects, the spatial frequency can be expressed in cycles/mrad.

The **optical cut-off frequency** is the spatial frequency beyond which an optical system cannot transmit information. It equals $2NA/\lambda$ or $1/(\lambda f/\#)$ for incoherent light.

In a perfect image, the MTF is unity at all frequencies. In a real image, contrast is degraded by both diffraction and aberrations. The OTF is equal to either the Fourier transform of the incoherent PSF or the autocorrelation of the OPD (pupil function). In lenses with large aberrations, the OTF can be negative, leading to **MTF bounce**, which represents a **phase reversal** of light and dark features (**spurious resolution**).

The **contrast transfer function (CTF)** uses a square-wave modulation to determine contrast and is specified in line pairs/mm. A line pair consists of one black line and one white line. MTF specifications often refer to line pairs/mm when cycles/mm is intended because square wave targets are easier to generate for measurement purposes.

Using MTF in Lens Design
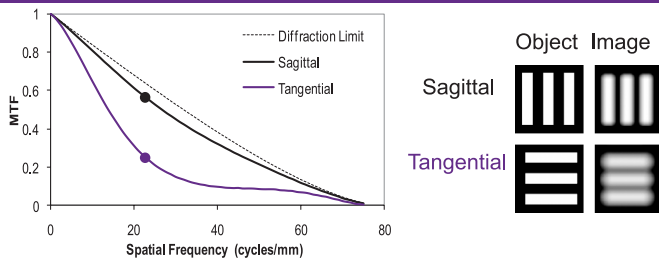
A complete MTF plot includes performance curves for several different field points and a **diffraction-limited** reference line. For off-axis field points, the MTF depends on the orientation of the "lines" in the object and is therefore plotted for both spoke-like radial modulation (**sagittal** S) and wheel-like azimuthal modulation (**tangential** T), where the modulation direction is orthogonal to the feature direction. The on-axis MTF is typically a single curve but will be separated into sagittal and tangential components in lenses with decentered or tilted elements.

> To aid in developing system performance budgets, the MTF can be indirectly linked to wave aberrations or OPD using the **Hopkins ratio**, which describes the ratio of the aberrated MTF to the diffraction-limited MTF as a function of RMS OPD.

Imaging performance is quantified by specifying an MTF value at a particular spatial frequency. For digital sensors, this frequency is usually associated with the sensor's **Nyquist frequency**. Although the MTF plots do not directly tell a designer which aberrations are present in the lens, MTF is useful for specifying system performance because it is easily measured and allows the cascading of MTFs using **linear system theory** (where the camera, atmosphere, lens, and other MTFs are multiplied to derive the overall system MTF). However, optimizing a lens directly for MTF is computationally intensive, and phase reversals in the MTF can cause local minima in the merit function and stagnated optimization.

Defocus

In an optical system, **focus** is the act of changing the relative position of the sensor with respect to the elements (usually to improve the sharpness of an image).

The **hyperfocal distance** is the object distance for a lens that maintains the depth of focus criteria at the image plane for all object distances between half the hyperfocal distance and infinity (without refocusing).

$$W(H,\rho,\theta) = W_{020}\rho^2$$

$$W_{020} = -\frac{n'u'^2}{2}\delta z$$
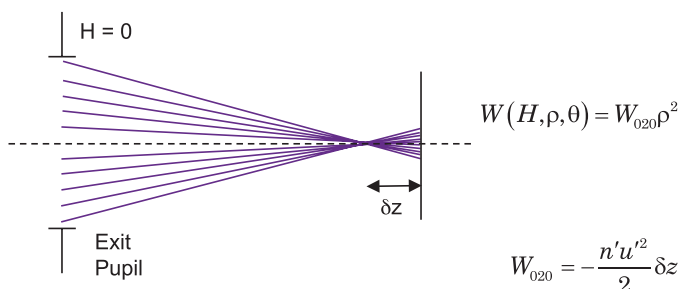
Defocus is the amount of aberration introduced when moving the image plane axially from the paraxial image plane and is directly proportional to the amount of movement δz. Defocus is used to balance the blur of other aberrations and improve the image quality of the system. It is not typically considered a "true" aberration.
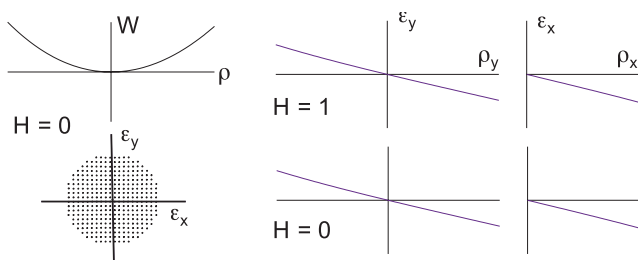
A wavefront plot for defocus shows a quadratic dependence on aperture coordinate ρ, whereas a transverse ray plot shows a linear dependence on ρ that has the same slope in both x and y and is independent of field coordinate H.

Wavefront Tilt

$$W(H,\rho,\theta) = W_{111}H\rho\cos\theta$$
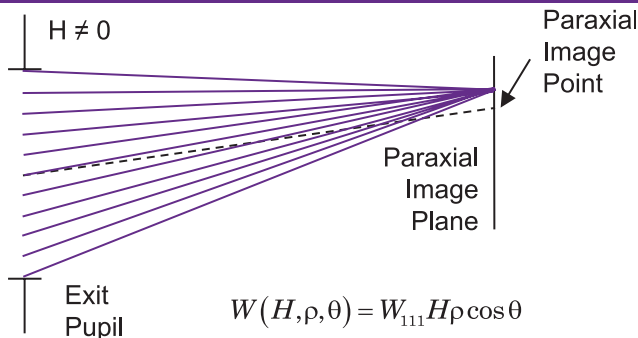
Wavefront tilt is the difference between the paraxial magnification and the actual magnification of the system. There is no tilt on axis ($H = 0$); for objects located along the y axis, there is no tilt in x. A wavefront plot for tilt shows a linear dependence on aperture coordinate ρ, whereas a transverse ray plot shows a constant dependence on ρ that increases in magnitude with field coordinate H. The spot diagram shows a perfect point image, but it is not located at the expected paraxial image point.
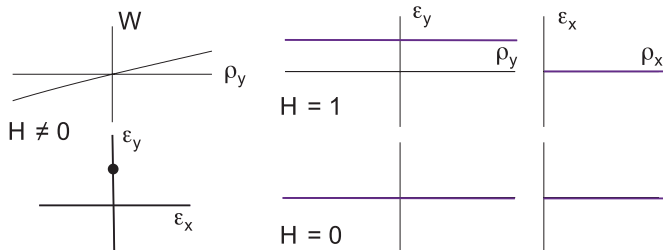

The **depth of focus** of a lens is the longitudinal shift of the image sensor that produces an image degradation that is still acceptable for the application. For example, the Rayleigh quarter-wave criterion for OPD results in a depth of focus equal to $\pm 2\lambda(f/\#)^2$ or $\pm\lambda/(2NA^2)$. Similarly, the **depth of field** of a lens is the allowable shift of the object position and is related to the depth of focus by the longitudinal magnification.

Spherical Aberration

Spherical aberration is a variation in focus position between the paraxial rays through the center of the lens aperture and those through the edge of the aperture.
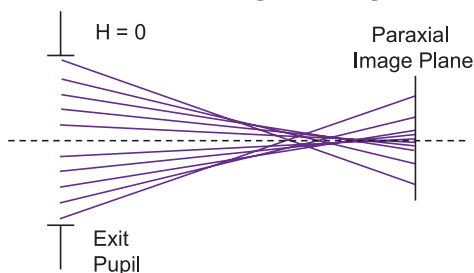
$$W(H,\rho,\theta) = W_{040}\rho^4$$

Spherical aberration is the only one of the five primary third-order aberrations to appear on axis ($H = 0$), and it has the same magnitude for any field coordinate H. A wavefront plot for spherical aberration shows a fourth-order dependence on aperture coordinate ρ, whereas a transverse ray plot shows a cubic dependence on ρ that is the same in both x and y.
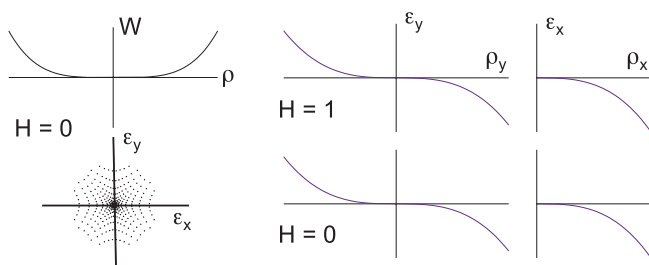
Spherical aberration is the first aberration to appear in most symmetrical lens tolerancing budgets. Because errors in the airspaces between lens components also result in spherical aberration, a **space adjust** procedure can be devised whereby adjustable spacers are used in the most sensitive airspaces of a design as **compensators** to fix the as-built performance.

Coma

For symmetrical optical systems, **coma** is an off-axis aberration where annular zones of the aperture have different magnifications.
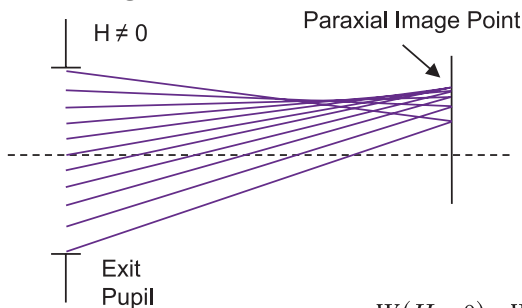
$$W(H,\rho,\theta) = W_{131} H \rho^3 \cos\theta$$

There is no coma on axis ($H = 0$), and for objects located along the y axis, there is no coma in x when ρ_y equals zero. A wavefront plot for coma shows a cubic dependence on aperture coordinate ρ, whereas a transverse ray plot shows a quadratic dependence on ρ that increases in magnitude with field coordinate H. The spot diagram shows that the resulting image of a point object looks like a comet.
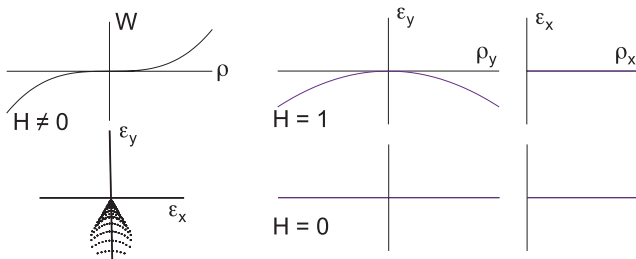
Coma is the first aberration to appear in most asymmetrical lens tolerancing budgets. For example, errors in the tilt or decenter of lens components result in coma. Correction schemes can be devised that include an adjustable "**push-around**" element as a compensator to fix the as-built performance.

Field Curvature

H ≠ 0

Perfect image lies on a curved focal surface

Paraxial Image Plane

Exit Pupil

$$W(H, \rho, \theta) = W_{220} H^2 \rho^2$$
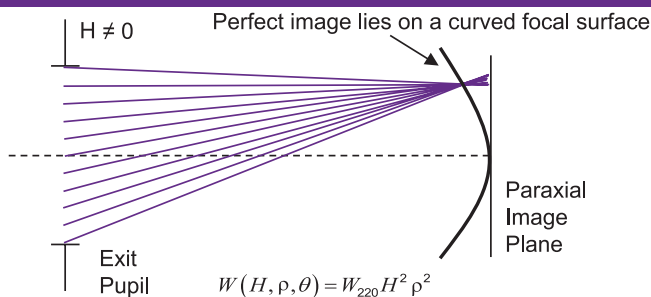
A symmetrical optical system with spherical surfaces naturally forms an image on a curved surface. **Field curvature** is an off-axis aberration that describes the departure of the image surface from a flat surface. Field curvature looks similar to defocus; however, there is no field curvature on axis ($H = 0$), and its magnitude increases with field coordinate H. A wavefront plot for field curvature shows a qu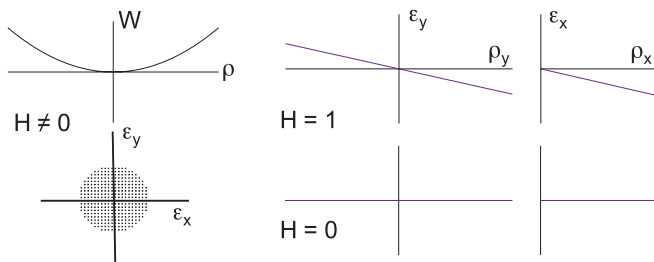adratic dependence on aperture coordinate ρ, whereas a transverse ray plot shows a linear dependence on ρ that has the same slope in both x and y.

W

ρ

H ≠ 0

ε_y

ε_x

ε_y

ρ_y

ε_x

ρ_x

H = 1

H = 0

A **field flattener** is a lens (usually negative) placed close to the image plane where the marginal ray height is small. These lenses are used to reduce the field curvature without significantly changing the image size or adding spherical aberration.

f/2, 7-deg HFOV

Field Flattener
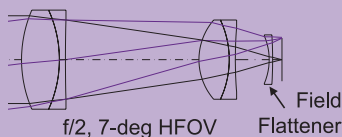
Petzval Curvature

The field curvature coefficient W_{220} can be separated into two components, where W_{222}^* is a term whose functional dependence is equivalent to the astigmatism coefficient (W_{222}), and W_{220p} is the portion of field curvature due to **Petzval curvature**.

$$W_{220} = \frac{1}{2}W_{222}^* + W_{220p}$$

The separation of field curvature into two distinct terms is useful during lens design, as the techniques for correcting the two aberrations are significantly different. The amount of astigmatism introduced by a lens element is a function of the power and shape of the lens and its distance from the aperture stop, whereas the amount of Petzval curvature introduced by that same element is purely a function of lens power and index of refraction. Petzval curvature is the only aberration completely independent of ray quantities (angles and heights); this makes Petzval curvature one of the most difficult aberrations to correct once the first-order power balance in the design has been established. The key to fixing the Petzval curvature (thereby reducing the overall field curvature) in optical design is to change the first-order power balance between positive and negative power.

In most positive-power refracting lenses, the Petzval surface will be curved toward the lens (known as an **inward-curving field**). If there is no astigmatism in the lens, the field curvature reduces to the Petzval curvature. If Petzval curvature cannot be corrected, overcorrecting astigmatism can be introduced to artificially flatten the tangential field, leaving the sagittal field to fall between the image plane and the Petzval surface, and yielding the smallest amount of residual field curvature.

Some optical systems use a curved image surface (e.g., the retina) to significantly reduce the effects of Petzval. Modern-day "curved" digital sensors have the potential to dramatically simplify lens systems by removing the need to correct Petzval in the design.

Astigmatism

Astigmatism is the aberration that causes a tangential fan of rays to focus at a different location than a sagittal ray fan. Fundamentally, astigmatism looks like field cur-

$$W(H,\rho,\theta) = W_{222}H^2\rho^2\cos^2\theta + \frac{1}{2}W_{222}{}^*H^2\rho^2$$

vature, but the magnitude of the aberration is different between *x* and *y*. In the wavefront expansion, astigmatism is defined in terms of a tangential (*y* axis) component only. However, a portion of the field curvature of a lens is also proportional to the astigmatism coefficient, adding a sagittal (*x* axis) component; this term is also included in the plots.
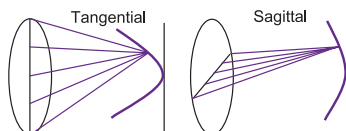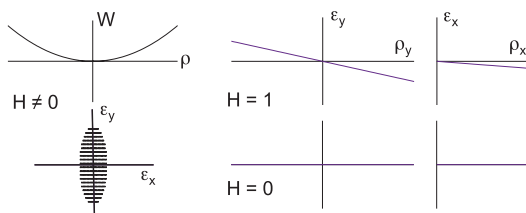
There is no astigmatism on axis ($H = 0$), and its magnitude increases with field coordinate *H*. A wavefront plot for astigmatism shows a quadratic dependence on aperture coordinate ρ, whereas a transverse ray plot shows a linear dependence on ρ that has a different slope in *x* and *y*. A spot diagram shows the image of a point to be an ellipse.

Design programs plot the longitudinal astigmatic **field curves** on a separate graph. When there is no astigmatism, the sagittal and tangential image surfaces coincide with each other and lie on the Petzval surface. If there is astigmatism, then the tangential image surface lies three times as far from the Petzval surface as the sagittal image surface.

Distortion

H ≠ 0

Paraxial Image Point

Paraxial Image Plane

Exit Pupil
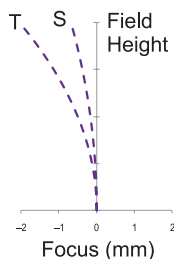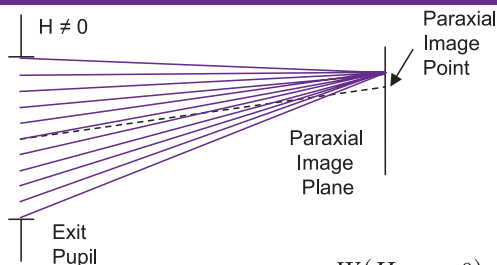
$$W(H, \rho, \cos\theta) = W_{311} H^3 \rho \cos\theta$$

Distortion is an aberration in which the magnification varies over the image height. Object points are imaged to perfect points with no blur, but straight lines in the object become curved lines in the image. Design programs do not include distortion as part of their wavefront or transverse ray plots. They instead plot distortion as either a percentage versus field coordinate H or as a deviation from a grid of points. The percent of distortion is usually calculated from the difference in height between a real and a paraxial chief ray for a given field point; however, a **centroid distortion** (using either the geometric spot or PSF centroid location instead) may be a more relevant evaluation for a particular lens/application.

H = 1

−1   −0.5   0   0.5   1
% Distortion

—— Paraxial Grid
—— Distorted Grid
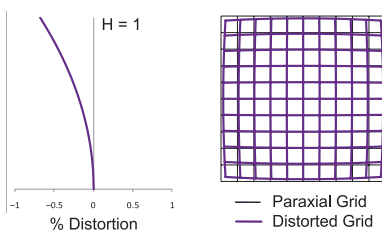
- **Pincushion distortion** occurs if the magnification increases toward the edge of the field.

- **Barrel distortion** occurs if the magnification decreases toward the edge of the field.

- **Keystone distortion** produces a trapezoidal image of a rectangular object and can result from nonparallel object and image planes.

Primary Color and Secondary Color

**Chromatic aberrations** arise because the refractive index of a lens material changes with wavelength (characterized by the material **dispersion**). Light from a single object point at different wavelengths is not

focused to the same point in the image (e.g., for a simple lens, blue light from an axial, white-light point source
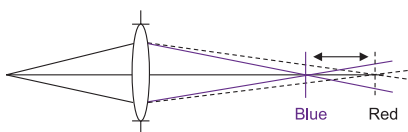
is focused before red light). For polychromatic lenses, performance evaluations such as transverse ray plots are computed for at least three wavelengths that span the system bandwidth (one at each end of the spectrum and one in the middle). For the visible spectrum, these wavelengths are typically C (red), d (yellow), and F (blue).

The difference between the red and blue focus is known as **primary color** (also referred to as **axial color**, **longitudinal color**, or **longitudinal chromatic aberration**). Similar to defocus, the transverse ray plot shows a linear dependence on ρ that has the same slope in both x and y and is independent of field coordinate H; but the slope of the line changes with wavelength.
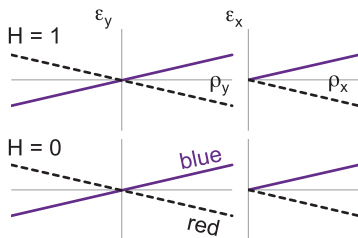
If the primary color of the lens has been corrected, the lens can still have **secondary color**, which is a difference between the common red/blue focus position and the yellow focus position (also indicated by a change in slope
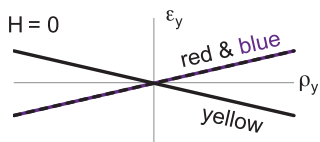
on the transverse ray plot). When the d line is also imaged to the same focus as the F and C lines, the system is said to be corrected for **secondary spectrum**.

Lateral Color and Spherochromatism

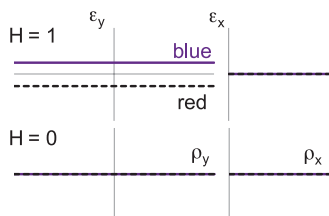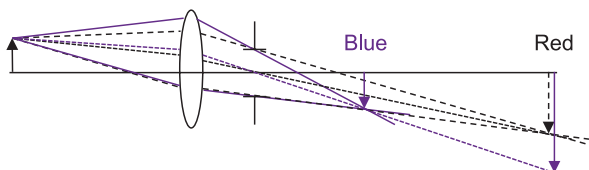**Lateral color** (or **transverse chromatic aberration**) is an aberration in which the magnification of the system is wavelength dependent (the image height varies with wavelength). For a simple visible lens with the aperture stop after the lens, the blue image appears to be larger than the red image.

There is no lateral color on axis ($H = 0$), and for objects located along the y axis, there is no lateral color in x. A transverse ray plot shows a constant dependence on ρ that increases in magnitude with field coordinate H (similar to tilt); however, the intercept value changes with wavelength. Unlike primary color, the magnitude of lateral color is highly dependent on the stop location.

Color correction equations developed for the visible spectrum (d, F, and C wavelengths) can be used outside of the visible spectrum if a custom Abbe numer V is defined for a material over the band of interest, with the indices at the long, short, and middle wavelengths.

Spherochromatism is a variation of spherical aberration with wavelength. In some lenses, there is very little spherochromatism. In others, the spherical aberration can switch sign (typically blue is **overcorrected** and red is **undercorrected**), and the spherochromatism must be corrected or balanced with axial color to reduce image blur.
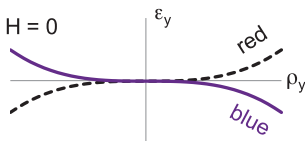
Higher-Order Aberrations

Lens systems with large apertures ($<f/4$) and wide fields of view (>50-deg FFOV) will show evidence of **higher-order (HO) aberrations**. The five third-order aberrations have fifth-order counterparts (e.g., **fifth-order spherical aberration** looks very similar to third-order spherical aberration, but its dependence on aperture coordinate ρ increases in order). There are also two "new" fifth-order aberrations: **elliptical coma** and **oblique spherical aberration**.

$$W\left(H,\rho,\cos\theta\right) = O(2) + O(4)$$

$+W_{060}\rho^6$	Fifth-order SA
$+W_{151}H\rho^5\cos\theta$	Fifth-order linear coma
$+W_{422}H^4\rho^2\cos^2\theta$	Fifth-order astigmatism
$+W_{420}H^4\rho^2$	Fifth-order field curvature
$+W_{511}H^5\rho\cos\theta$	Fifth-order distortion
$+W_{240}H^2\rho^4$	Sagittal Oblique SA
$+W_{242}H^2\rho^4\cos^2\theta$	Tangential Oblique SA
$+W_{331}H^3\rho^3\cos\theta$	Cubic coma ⎤ Elliptical
$+W_{333}H^3\rho^3\cos^3\theta$	Line coma ⎦ coma

+ Higher-order terms

During optimization, the best performance is typically achieved by balancing low-order aberrations with high-order aberrations. The residual errors then will show an inflection in the transverse ray, distortion, or field curves, indicating a sign change between the low-order and high-order aberration contributions.
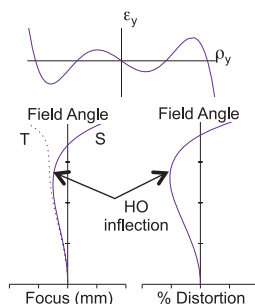
Intrinsic and Induced Aberrations

Intrinsic aberrations are aberrations associated with a single surface in an optical system and can be calculated solely from paraxial ray angles and heights at a given surface. In contrast, **induced aberrations** are aberrations generated from the interaction between aberrated rays coming into a surface and the intrinsic aberrations of that surface.

A simple example of induced spherochromatism can be constructed by placing a concave spherical mirror in the converging cone created by a positive-powered singlet. The lens sends uncorrected axial color into the mirror surface so that the margina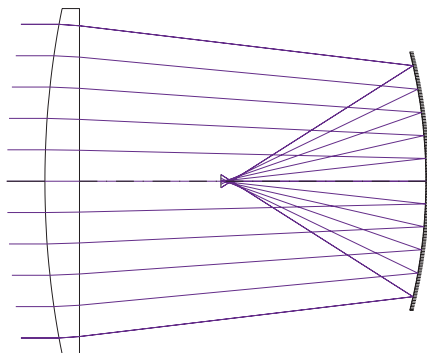l ray height at the surface varies with wavelength; the axial color interacts with the mirror's intrinsic spherical aberration resulting in a different amount of spherical aberration for each wavelength. The mirror itself is not dispersive, so the aberration is not intrinsic.

Intrinsic and induced aberrations can be used to offset each other to improve performance. The functional form of an induced aberration can be identical to an intrinsic aberration, but it is usually harder to correct the induced aberrations in a design. Complex, well-corrected designs are often limited by higher-order induced aberrations (specifically, higher-order Petzval curvature and sagittal oblique spherical aberration).

Selecting a Design Form: Refractive

Aperture and **field size** are the two key system specifications that determine the difficulty/complexity of a design and can therefore be used to identify an appropriate starting design form. If the aperture or field size of a design form is increased outside of its working region, the performance drops off; if the aperture or field size is decreased, the design form is likely to meet the performance requirement but is more complex/costly than necessary. The classic design-form regions indicated below are approximate and do not necessarily represent diffraction-limited performance. Most assume spherical surfaces and are valid over the visible band. Modern designs build on these basic forms, often incorporating aspheres to reduce package size, improve performance, and/or increase the aperture or field size.
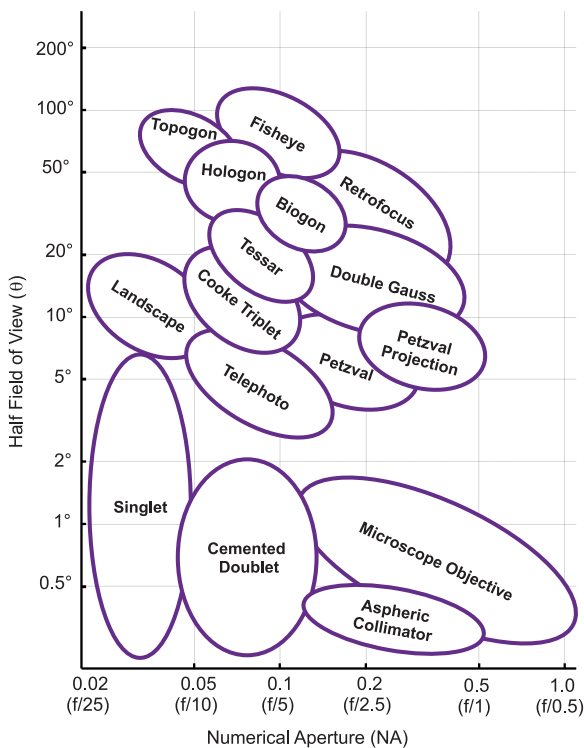
Selecting a Design Form: Reflective

Similar to refractive design form selection, aperture and field size are the two key metrics for choosing a reflective design form. It is important to first select a form that can support the FOV requirements. Then the aperture specification is a strong driver for the number of mirrors and their asphericity; aspheres are effective at correcting spherical aberration when increasing the aperture but are less effective at increasing the field coverage. The system size requirements will often filter out most of the candidate design forms. For example, if the envelope diameter must be less than 2–3× the EPD, it is unlikely that a system with a negative primary mirror that performs well over a wide field can be used. If an accessible field stop and/or cold stop is required, this implies that a reimaging system is needed and excludes non-reimaging forms. Furthermore, if a Lyot stop is needed with high-quality pupil imaging, more complexity in the mirrors (e.g., tilts and/or higher-order aspherics) might be necessary.
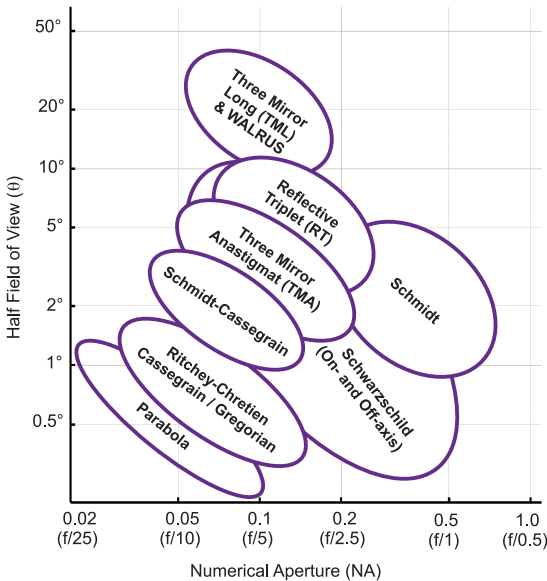
Singlets
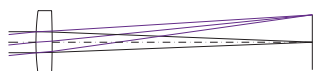
f/15, 5-deg HFOV

A **singlet** is a lens used without any other powered elements; the stop is typically located at the lens. Singlets have relatively poor image quality, limited by large chromatic aberrations and field curvature. They are often used as a simple magnifying glass or photon collector.
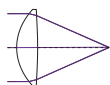
An **aspheric collimator** is a singlet with at least one aspheric surface. The aspheric surface allows for an increased numerical aperture and increased resolution

f/1.25, 0.25-deg HFOV

over a very narrow wavelength range and small field. These lenses are commonly found in laser collimating applications such as fiber coupling, laser diodes, optical data storage, and barcode scanners.

Landscape photography requires wide-FOV lenses that capture as much of the scenery as possible. Shifting the stop in a singlet away from the lens increases its FOV and results in a simple **landscape lens**. As the stop is moved, the lens shape changes and is generally bent around the stop to reduce the ray angles of incidence and the off-axis aberrations. The stop shift also introduces astigmatism that helps to flatten the tangential field curvature. A single-element landscape lens still has spherical aberration and chromatic aberrations and is therefore restricted to a small NA. For better correction, the singlet may be aspherized or compounded into a doublet.
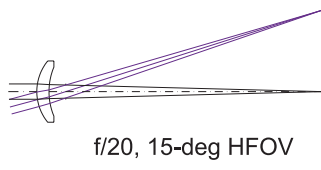
f/20, 15-deg HFOV

The farther the aperture stop is shifted from the lens, the larger the lens diameter. Disposable "box" cameras generally use simple landscape lenses.

Achromatic Doublets
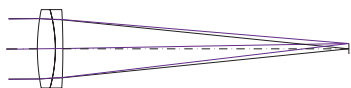
f/5, 1-deg HFOV

An **achromatic doublet** or **achromat** is composed of two singlets designed in combination to correct primary color at two wavelengths (typically red and blue). The individual element powers of the doublet are determined by the total power and the individual Abbe numbers of the materials. For standard refracting materials with positive Abbe numbers, the two elements will have opposite powers, and each has stronger absolute power than the total doublet power. The shape of the elements is chosen to minimize spherical aberration and coma.

The on-axis spot size of an $f/5$ visible-band achromat is ~25 times smaller than a singlet of the same aperture.

Doublet elements can be either cemented together or separated by an airspace. The **cemented doublet** is often preferred for ease of fabrication and alignment. A **Fraunhofer doublet** is a positive crown followed by a negative flint, whereas a **Steinheil doublet** is a flint-first doublet. The Fraunhofer form has a shallower curvature at the cemented interface and is usually the preferred design form.

Achromats are limited in field to less than a few degrees by residual Petzval and in $f/\#$ to $f/4$ or larger by higher-order spherical aberration. Fast achromats ($<f/4$) are rare because the V ratio that determines element power is fundamentally limited by existing glass choices. Common doublet applications include broadband collimating lenses and telescope objectives. Doublets are also critical building blocks for complex broadband lenses.

Lenses corrected for axial color at three and four wavelengths are called **apochromats** and **super-achromats**, respectively, and usually require more than two elements/materials.

Airspaced Doublets

The speed of a **Fraunhofer doublet** can be increased to ~$f/2.5$ by separating the cemented surface to create an **airspaced doublet**. The extra curvature and airspace are used to control higher-order spherical aberration and sphe-

Fraunhofer   Gauss

rochromatism. However, the ray angles of incidence in the airspace between the two elements can become very large, making the alignment tolerances of the two elements very tight. A second type of airspaced doublet is a **Gauss doublet**, which has much steeper curves and significantly larger higher-order aberrations. As a result, Gauss doublets are not as fast as Fraunhofer doublets and are often found in stop-symmetric anastigmatic lenses.
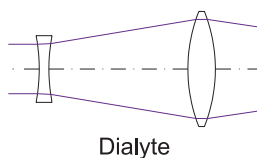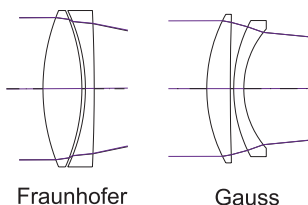
Dialyte

Two separated positive and negative elements (or groups of elements) with a relatively large air gap between them is known as a **dialyte**. Although the performance of a simple two-element dialyte is usually insufficient (primarily due to its inability to correct lateral color), it is the basis for more-complex **telephoto** (positive/negative) and **reverse telephoto** (negative/positive) multi-element designs.

A **Schupmann lens** is a special airspaced doublet that is corrected for primary color but uses the same material in both elements. Because the lens has a virtual image, it is seldom used alone but rather is often part of more-complex lens systems.

Schupmann
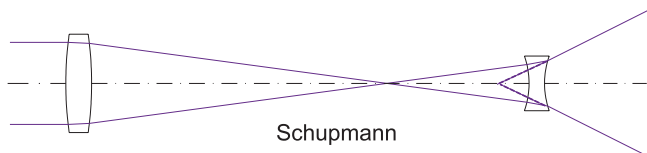
Cooke Triplet

The **Cooke triplet** is the simplest design form with enough degrees of freedom to correct all first- and third-

f/6, 10-deg HFOV

order aberrations and hold first-order imaging constraints such as EFL and BFL. The outer two elements are positive crowns, and the middle element is a negative flint with the aperture stop located on either side of the middle element. The symmetry of the design helps control coma, distortion, and lateral color. Because it is only corrected to third order, the Cooke triplet is limited by higher-order aberrations and is thus capable of relatively slow apertures ($\sim f/6$) and moderate fields (10–15 deg). Up to 50% vignetting may be needed in systems with larger fields.
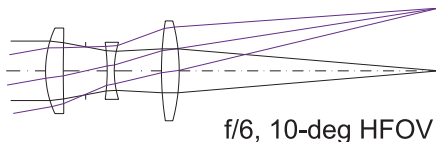
The Cooke triplet is the most basic form of an **anastigmat lens**. Strictly speaking, the term anastigmat means "free of astigmatism;" however, it is also used to describe optical systems with reduced astigmatism and relatively flat fields. In these designs, the field is flattened by separating positive and negative power along the optical axis (between surfaces, elements, or even multi-element groups of components). **PLAN microscope objectives** use this principal to flatten the field curvature.

Many **photographic objective** design forms are derived from the Cooke triplet by **splitting**, **compounding**, or otherwise modifying elements of the basic triplet. For example, in the **Tessar** lens, the last element of a Cooke triplet is compounded into a doublet. This design allows for speeds up to $f/4$ and/or larger fields of view than the basic triplet with reduced vignetting.

f/6, 20-deg HFOV
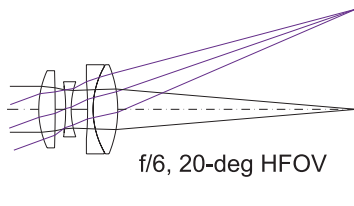
Double Gauss

The **double Gauss** (or **Biotar** or **Planar**) is a powerful anastigmat design form used in a wide range of applications, including fixed-focus camera objectives, projection lenses,

f/2, 20-deg HFOV

and high-resolution microscope objectives. The classic six-element design is nearly symmetric about the stop (reducing coma, distortion, and lateral color), with outer positive singlets and inner cemented thick-meniscus doublets (to minimize Petzval). Astigmatism is controlled by adjusting the distance of the elements from the stop, as this separation has little effect on the other aberrations. Design improvements to the basic form have been studied extensively in the literature; allowing the system to depart from symmetry and/or adding elements (by splitting or compounding) allows one to achieve speeds up to $f/1$ or FFOVs larger than 50 deg.
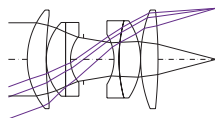
Fast anastigmats with small fields tend to have relatively short vertex lengths, whereas slow-speed wide-angle lenses tend to have longer vertex lengths. In Cooke triplets the vertex length can be controlled by glass choice (particularly the difference in V between positive and negative elements).

Many other lens forms use stop symmetry and power separation to achieve wide fields of view. The simplest is the **Hypergon**; two identical meniscus elements symmetric about the stop achieving an impressive 135-deg FFOV. Petzval is reduced by the separation of positive and negative power in the elements themselves where each element has a positive and a negative surface with radii differing by less than 7%. However, Hypergons are slow ($\sim f/20$) as there is no remaining degree of freedom to correct the spherical aberration. More-complex designs (e.g., **Topogon**, **Hologon**, and **Biogon**) add elements to increase their speed.
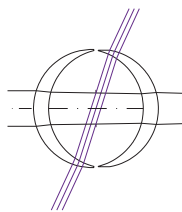
f/20, 65-deg HFOV

Petzval Lens

The **Petzval** or **Petzval portrait lens** is a design form useful for high-speed applications that require high-quality imaging over relatively small fields (e.g., aerial reconnaissance). Originally designed for indoor photography, Petzval lenses are an order of magnitude faster than landscape lenses. The Petzval lens is the basis of many microscope objectives where aplanatic elements are added to the short conjugate for increased NA.

f/2.5, 7.5-deg HFOV
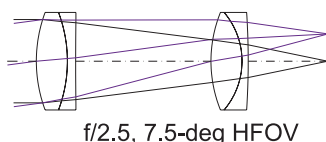
Unlike many of the other basic design forms, the Petzval lens has very little symmetry. The basic layout uses two separated, positively powered achromats with the stop at or near the front lens. Negative astigmatism from the first doublet is balanced with positive astigmatism from the second doublet. Because there are no lens groups with net negative power, the design form is fundamentally limited by Petzval, restricting its field size. Decreasing the airspace between the two doublets can help reduce the Petzval at the expense of increased astigmatism.

Given the focal length f of a Petzval lens, the first-order thin lens layout can be calculated as follows:

- Front element focal length: $2f$
- Rear element focal length: f
- Element separation: f
- Back focal length: $f/2$

Improvements to the basic Petzval lens include adding a strong negative **field lens** near the image plane to correct Petzval and increase its field (although this significantly decreases the working distance).

f/1.5, 7.5-deg HFOV
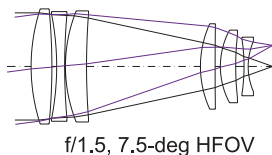
The Petzval lens is also the base design form of many extremely fast ($\sim f/1$) lenses, including very high-speed, 6–8-element **Petzval projection lenses**.

Telephoto Lenses
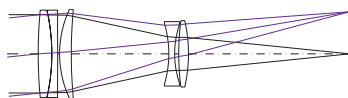
f/5, 6-deg HFOV

A basic **telephoto lens** consists of two lens groups—a positive group followed by a negative group. The stop is at or near the front group. The front principal plane is pushed out in front of the lens so that the EFL of a telephoto lens is greater than its physical length (measured from the first lens surface to the image plane). The ratio of the lens length to EFL is called the **telephoto ratio**. Typical telephoto ratios range from 0.6 to 0.9; the smaller the ratio is, the more compact the system. The asymmetry of the design form limits its field (<10-deg HFOV). Moderate-speed (~f/6) lenses can be built with two cemented doublets, whereas faster lenses can require the doublets to be air spaced and/or an element to be added to the first or second lens group.

The focal lengths of any two-element thin lens system can be calculated given the system focal length f, their separation d, and the back focal length *BFL*:

$$f_a = \frac{df}{f - BFL} \qquad f_b = \frac{dBFL}{f - BFL - d}$$

For zero-Petzval sum:

$$f_a = -f_b = f - BFL \qquad d = \frac{(f - BFL)^2}{f}$$

Both lens groups are usually separately achromatized as primary color and secondary spectrum limit the image quality of telephoto lenses. The telephoto ratio cannot be arbitrarily reduced without increasing Petzval (which limits the field coverage of the design). Faster lenses and smaller telephoto ratios require more-complex designs. Long-focal-length telephotos are particularly sensitive to small changes in object distance, and if used for photography, require an additional **floating element** that adjusts the focus as the object position is changed.

Retrofocus and Wide-Angle Lenses

f/4, 35-deg HFOV
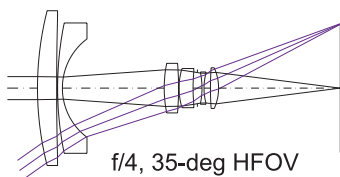
A **retrofocus lens** (or **reverse telephoto**) is composed of a negative lens group followed by a positive lens group producing a BFL much greater than its EFL (the opposite of a telephoto lens). This form is extremely useful for short-focal-length lenses that need a long working distance (e.g., for the placement of an SLR viewfinder mirror or a color-separation prism). This design form is also the basis for most **wide-angle lenses** covering large fields. Extremely wide-angle or **fish-eye lenses** can cover more than 180-deg full field.

f/4, 80-deg HFOV
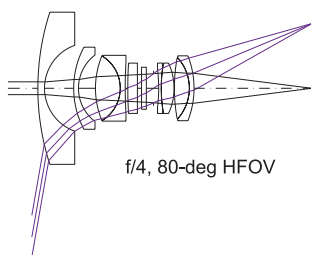
Retrofocus lenses are large in both diameter and length (can be 10–20× their focal length). Many telephoto design techniques are also valid for retrofocus lenses (e.g., component achromatization and Petzval reduction). A large airspace between the component groups helps reduce their individual power and aids in correction. The lack of symmetry makes correcting secondary lateral color and chromatic distortion a challenge in these designs. A conic asphere on the inner surface of the front element can help the overall correction of a wide-angle lens.

Wide-angle lenses suffer from significant third-order barrel distortion that is difficult to correct and is often balanced with a fifth-order pincushion distortion. Fish-eye lenses have an inherently large barrel distortion that is not necessarily considered to be an aberration but rather a necessity for imaging a 180-deg hemispherical scene onto a flat image plane. The standard $h = f \tan\theta$ definition of focal length no longer holds and may be substituted with $h = f\theta$ or $h = f \sin\theta$ distortion constraint. The barrel distortion also offsets the illumination fall-off that results from the cosine-fourth-power rule at large field angles.

Refractive versus Reflective Systems

Optical systems can be **refractive** (lenses), **reflective** (mirrors), or **catadioptric** (both lenses and mirrors). Because many refractive design forms have analogous reflective forms (e.g., the telephoto and the Cassegrain, the retrofocus and the Schwarzchild, or the Cooke triplet and the reflective triplet), it is useful to understand the advantages and disadvantages of one type over the other.

Advantages of reflective systems:

- No chromatic aberration (a single system can then be "shared" among many spectral bands/channels)
- Spherical aberration for a spherical mirror is 1/8 of that for a lens with a glass index 1.5 and an equivalent *f*/#
- Opposite sign of Petzval compared to refractive lenses
- Better performance over a wider thermal range
- Fewer surfaces needed
- Integrated mounting and alignment features
- Elements can be independently shaped to reduce weight
- Design can be folded to meet packaging constraints.

Disadvantages of reflective systems:

- Limited FOV
- Aspheres needed for high-quality performance
- More sensitive to irregularity and alignment errors
- Per-element fabrication cost and schedule
- Stray light is harder to manage
- Complex off-axis configurations needed to get mirrors in the light path without obscuring the incoming beam.

In very wide-angle lenses, rays perpendicular to the optical axis can fail to trace because their direction cosines are zero. Imaging the stop through the front group of negative elements also yields an anamorphic entrance pupil that needs to move forward, shift laterally, and tilt to get light into the lens as the field angle increases. In software, ray aiming or a "wide-angle mode" (uses real chief rays versus paraxial rays) is needed to overcome these ray failures.

Obscurations

In reflective systems, **obscurations** are objects that block a portion of the incident light bundle. For example, the image from a single on-axis parabolic mirror is located directly in the incoming beam path. If a detector is placed at the image plane, it will obscure the incoming beam. Common on-axis astronomical telescopes (e.g., Cassegrain or Gregorian) have secondary mirror obscurations. Although obscurations help reduce package volume for systems with space constraints, they cause an overall transmission or throughput loss. The stray light from obscurations also requires critical attention (especially in the infrared where the thermal effects of blackbody radiation from obscuring features must be considered).

It is important to clarify if the **obscuration percentage** is specified as a **linear obscuration** (% of the aperture radius) or an **area obscuration** (% of the aperture area); linear specifications are more common. A 20% linear obscuration is ~4% by area.

Obscurations modify the diffraction pattern of an object by moving energy from the central Airy disc to the outer rings. **Central obscurations** (blocking light in the center of the pupil) are the most common. Central obscurations primarily reduce the image contrast (MTF) at the mid-spatial frequencies. The drop in MTF becomes noticeable around 15% obscuration (measured linearly) and increases as the size of the obscuration increases. A second common type of obscuration is a structural support mount or **spider** used to hold mirrors in place. Each spider vane produces a diffraction spike in the PSF at a right angle to the direction of the vane. Obscurations can also induce vignetting and, if near the image plane, can directly alter the shape of the final image (e.g., Hubble Space telescope WFPC).
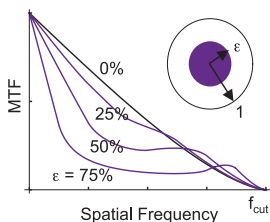
Newtonian and Cassegrain
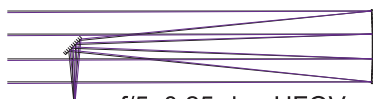
f/5, 0.25-deg HFOV

A single parabolic mirror used with a fold mirror is called a **Newtonian telescope**. It has zero spherical aberration on axis for an infinitely distant object but is quickly limited off-axis by coma. A **Cassegrain telescope** adds a negatively powered hyperbolic secondary mirror to a parabolic primary. Both mirrors share a common focus point, forming a mathematically perfect, aberration-free or **stigmatic** image of an infinite on-axis object. The field is limited by coma and astigmatism to less than a degree, but the compact package and small telephoto ratio make the form extremely attractive for many portable applications. Shifting the mirrors axially from stigmatic alignment balances coma with spherical, improving the performance in the field.
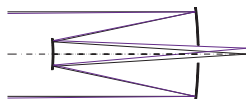
f/8, 0.5-deg HFOV

In a first-order layout, the mirror curvatures of any two-mirror system are given by

$$c_1 = \frac{BFL - f}{2df} \quad c_2 = \frac{BFL + d - f}{2dBFL}$$

where f is the effective focal length, d is the separation between the two mirrors, and BFL is the distance from the second mirror to the image.

The **Ritchey–Chrétien telescope** uses a hyperbolic primary and hyperbolic secondary to correct spherical and coma exactly to third order. The field coverage is slightly larger than that of a Cassegrain; however, the primary mirror can no longer be tested with an on-axis diverging spherical beam on the finite-conjugate side, requiring the use of a **null lens**. Refractive **corrector lenses** can also be inserted to improve field performance, but they add chromatic aberration, limiting the telescope's spectral region.

Gregorian and Schwarzschild

A **Gregorian telescope** has a parabolic primary and positively powered elliptical secondary. By aligning the focus of each conic sec-tion, a perfect on-axis stigmatic im-

f/8, 0.5-deg HFOV

age is formed. Many early telescopes were Gregorian; his-torically, positive concave mirrors were easier to fabricate. However, the Gregorian with two positive mirrors has more field curvature than a Cassegrain. Modern mirror-fabrication techniques have made the Cassegrain more common. The performance of a Gregorian telescope can be improved by departing from the stigmatic imaging condi-tion and changing the primary mirror to an ellipse.
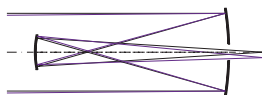
The 200″ **Hale Telescope** at Mt. Palomar is a Cassegrain; the **Hubble Space Telescope (HST)** is a Ritchey–Chrétien; and the **Large Binocular Telescope (LBT)** in Arizona is a Gregorian design.

A **Schwarzschild objective** has a small negative spherical primary and a large positive spherical secondary. If the mirrors are concentric with the stop located at their common center of

f/2, 1-deg HFOV

curvature, the system can be exactly corrected for third-order spherical, coma, and astigmatism, and the image lies on a curved surface whose radius equals the focal length. The mirror separation d and the mirror curvatures c_1 and c_2 are given by
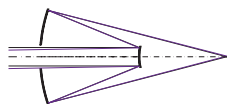
$$d = 2f \qquad c_1 = \left(\sqrt{5} - 1\right)f \qquad c_2 = \left(\sqrt{5} + 1\right)f$$

for a system focal length f. The size of the secondary mirror and large obscuration ratio (>40%) dramatically limit the use of a Schwarzschild as a telescope objective, but its low $f/\#$ and long working distance make it an attractive form for reflective microscope objectives. In practice, the mirrors may be aspherized and/or lenses added to the design to increase the NA and field coverage.

Catadioptric Telescope Objectives

**Catadioptric telescope objectives** use weak refractive elements to correct the aberration of high-power spherical mirrors without adding significant chromatic aberration.
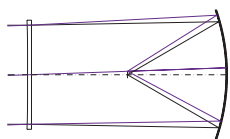
The simplest example is a **Mangin mirror** or back-surface reflective element with two powered surfaces. In a **Schmidt telescope**, the aperture stop is placed at the center of curvature of a spherical mirror, eliminating coma, astigmatism, and distortion. A thin corrector plate with one aspheric surface is then located at the stop, correcting all orders of spherical while adding only a small amount of chromatic aberration. The final image plane is curved with a radius equal to the focal length. A **field lens** can be added near the image plane to flatten the image surface, if needed. In practice, **ghost reflections** from the corrector plate can be a problem. Spherical airspaced doublet or triplet corrector plates yield **Houghton** or **Buchroeder** systems with improved achromatization without the use of aspheres.

f/1, 2-deg HFOV
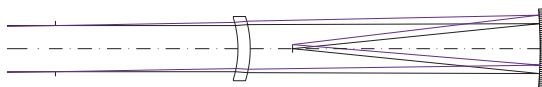
f/5, 1-deg HFOV

The **Maksutov** or **Bouwers telescope** has a spherical primary mirror and a weak-meniscus, concentric spherical corrector plate (historically easier to fabricate than the aspheric Schmidt plate). Nominally, all three surfaces are concentric to the stop; however, the corrector plate can reduce axial color with a slight departure from concentricity at the expense of slightly increased spherical. Both the Schmidt and Maksutov can be combined with other telescope objective forms. For example, the **Schmidt–Cassegrain** and **Maksutov–Cassegrain** are common commercial telescopes for amateur astronomy that combine the corrector optics with a Cassegrain system.

Unobscured Systems: Aperture Clearance
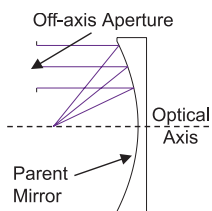
One method to eliminate the obscuration in a reflective system involves shifting the on-axis beam so that it is no longer centered on the optical axis (usually at the expense of increased package size). For example, in an **off-axis parabola (OAP)**, the aperture stop is simply decentered until the image is clear of the beam. It is important to note that the object, image, and mirrors still share a common optical axis in the **unobscured system**. The aperture may be made noncircular to keep the package size small, but this will change the shape of the PSF.

Tips for designing complex multi-mirror designs:

- The design process typically starts with definition of the size and speed of the primary mirror.
- On-axis conic sections, when used off-axis, suffer mainly from coma, and suffer from spherical aberration at nonideal conjugates.
- An analysis of the obscured system with nonapertured **parent mirrors** helps determine the aberration limitations of the unobscured design.
- Many unobscured systems use both aperture and field clearance to eliminate obscurations.

Unobscured two-mirror **afocal systems** can be constructed using two off-axis parabolas. These systems have perfect on-axis imagery

2×, 0.5-deg HFOV

and no coma or astigmatism (Petzval dominates the field). When set up as two positive confocal parabolas with an internal image plane, the exit pupil is real and accessible. A **Mersenne** configuration with a negative mirror has less Petzval and a smaller package size, but a virtual exit pupil. Both configurations are typically used as attachments to increase the aperture and focal length of a system or, conversely, shorten its focal length and increase its FOV.
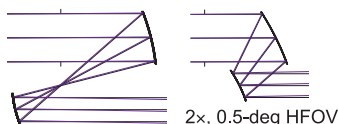
Unobscured Systems: Field Clearance

For unobscured infinite-conjugate systems, **field clearance** is achieved by tilting the entire system with respect to the incoming axial beam (this is equivalent to using an angular field definition that is not centered on the optical axis). For unobscured finite-conjugate systems, field clearance is achieved by eliminating the on-axis field point and using the system "off-axis in field," where the center of the field is decentered until the beam is clear of all obscurations (often at the expense of performance and packaging volume). Although it is often not obvious, the object, image, and mirrors share a common optical axis. If needed, the image plane and/or mirrors can be slightly tilted or decentered to improve performance; however, this change significantly increases the alignment complexity. As most detectors are rectangular, the smaller dimension of field is located in the plane of the field decenter to minimize package size. It is important to make sure that the detector is oriented correctly in the system, as the mirrors are optimized and sized for one image orientation only.

The **Offner relay** is a classic example of an unobscured finite-conjugate all-reflective system. It is a 1:1 system composed of two concentric
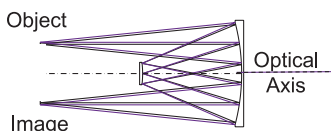

spherical mirrors with the object (and image) plane at their common center of curvature. The stop is located at the secondary mirror, and the system is doubly telecentric. The Offner design form eliminates all third-order aberrations exactly and is limited by higher-order astigmatism (but can be improved by deviating from perfect concentricity). Early high-performance, 1:1 ring-field lithography systems were based on the Offner design form. This type of design is also the preferred form for many spectrometers, where the secondary mirror is replaced by a reflective grating.

Three-Mirror Anastigmat
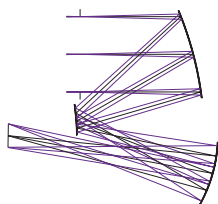

f/6, 2-deg × 1.5-deg HFOV

The Cassegrain telescope is a promising design form, as it can fit a long focal length in a short package while balancing Petzval with a positive mirror and a negative mirror. However, similar to many two-mirror design forms, the field coverage of a Cassegrain is limited to ~1 deg by astigmatism. Adding a third mirror to a Cassegrain provides the degrees of freedom necessary to correct the astigmatism across the field. The result is a **three-mirror anastigmat (TMA)** design useful for applications with HFOVs of 2–5 deg. In a focal TMA design, the tertiary mirror is an ellipse; in an afocal TMA design, the tertiary mirror is a parabola. Both obscured and unobscured (usually using aperture clearance) designs are possible. The key feature and main advantage of the TMA over other three-mirror design forms is that it is a **reimaging design** with an accessible internal image plane and an accessible exit pupil. This allows the placement of field stops,

Lyot stops, and cold stops (for infrared systems) for improved **stray light suppression**. Other TMA variations include designs with the internal focus after the primary rather than after the secondary, or designs with remote exit pupils for test setups.
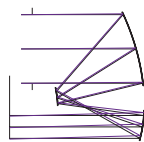
3×, 1-deg × 0.5-deg HFOV

Adding a fourth "corrector" mirror near the internal focus of a TMA improves the wavefront and reduces the distortion within a given FOV. The additional mirror also helps extend the field size in the direction of the mirror offsets to allow a square or circular FOV. The corrector mirror can be oriented coaxially with the other three mirrors or highly tilted to fold the design into a ball-shaped enclosure for pod-mounted applications.

Reflective Triplet
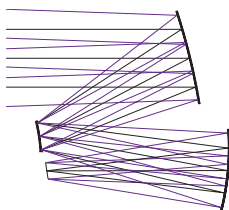
f/4, 2-deg × 2-deg HFOV

The **non-reimaging TMA** or **reflective triplet (RT)** is the reflective equivalent of the refractive Cooke triplet. The stop is at or near a negative secondary mirror with a positively powered mirror on each side of the stop. All three mirrors are conics and nominally share a common optical axis (the mirrors may be slightly decentered or tilted to improve performance). Because there is no intermediate image, the RT design form can support a larger field than the reimaging TMA design form; however, an RT design is not appropriate for thermally cooled systems, as the exit pupil is virtual and therefore inaccessible for the placement of a cold stop. The RT design form is recommended for moderate-FOV (up to 10-deg HFOV) applications that do not require reimaging. RTs perform best with the aperture stop at secondary (distortion is easier to correct due to the symmetry), but the design form also supports an aperture stop in front of the system, if needed. The beam clearance in an unobscured RT design is provided through both an off-axis aperture and an off-axis FOV. Systems that are off-axis in field will tilt and decenter the image plane to compensate for residual field curvature.

Distortion in an optical system is defined as the difference in image height between a real ray and a paraxial ray. For systems with centered FOVs the specification can be given by a single number at the edge of the field (e.g., 4% distortion). However, for unobscured designs with off-axis FOVs, the paraxial region may not even be in the FOV, and a single specification is not adequate. In this situation, a series of new terms such as "**smile**," **tow**/**keystone**, **expansion**, and **anamorphism** are needed to describe and quantify the distortion.
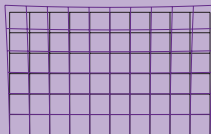
Wide-Field Reflective Design Forms

Wide-angle mirror de-
signs with more than
a 10-deg HFOV use a
design principle simi-
lar to wide-angle in-
verted telephoto refrac-
tive lenses. A front neg-

f/4, 10-deg × 10-deg HFOV
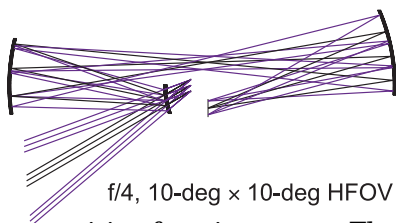
ative mirror is followed by a positive focusing group. The
aperture stop is located inside the lens (usually near the
positive focusing group). However, wide-angle reflective
designs tend to be large (many times larger than the en-
trance pupil diameter). The simplest example is an **off-
axis Schwarzschild**, which can also be converted to a
reimaging form for stray light suppression by adding a
concave relay mirror on the other side of focus in the
same way that a Cassegrain is converted to a TMA. The
additional mirror extends the FOV beyond what a stan-
dard Schwarzschild can achieve (e.g., up to 20 × 20 deg or
30 × 30 deg).

Afocal versions of 3- and 4-mirror reflective designs are
often used as beam reducers for optical systems needing
both large apertures (to collect light from distant objects)
and limited diameter filters, scan mirrors, diffraction
gratings, or beamsplitters that need to be placed in a
collimated beam.

Extremely large (by reflective standards > 20 deg), wide-
angle reflective designs are based on the design concept
that when the entrance pupil is at one focus of a
negative hyperbolic primary mirror, the reflection is free
of astigmatism. An aspheric secondary mirror is placed at
the aperture stop to control spherical aberration, and a
concave tertiary mirror is used to form the image. This
configuration has been referred to as the **three mirror
long (TML)**, as opposed to the **three mirror compact**,
another name for the reflective triplet. Variations on
the TML concept include forms with an aerial aperture
stop or **WALRUS**. The TML family provides the widest
field coverage of the reflective design forms. The main
disadvantage of this form is its very large envelope.

Zoom Lens Fundamentals

A **zoom lens** is a lens whose focal length can be changed either continuously or in discrete steps. **Continuous zooms** use **zoom groups** with fixed power that move along the optical axis. **Discrete zooms** laterally switch different fixed-power groups into (or out of) the optical path. Some zoom lenses (e.g., liquid lenses) can dynamically alter element powers without motion. "True" zoom lenses have a fixed image-plane location. **Varifocal**

lenses change focal length but allow the image plane to move, requiring refocus.

Moving Groups

Telephoto

Wide Angle

The **zoom ratio** is the ratio of the two extreme focal lengths (or, in the case of a finite conjugate zoom, the two extreme magnifications). For example, a 25–200-mm zoom lens is an 8× zoom. Zoom lenses are now standard for most consumer cameras, including "point and shoot" cameras, SLR cameras, and video camcorders, where the primary trade-off is large zoom range versus compact package size.

The $f/\#$ of most modern camera zoom lenses varies across the zoom range. At long focal lengths the aberrations tend to be larger, and the $f/\#$ is slower. A smaller aperture also reduces the physical size of the outer lens components. Distortion and lateral color will often change sign through zoom. The relative illumination and vignetting can change significantly through zoom.
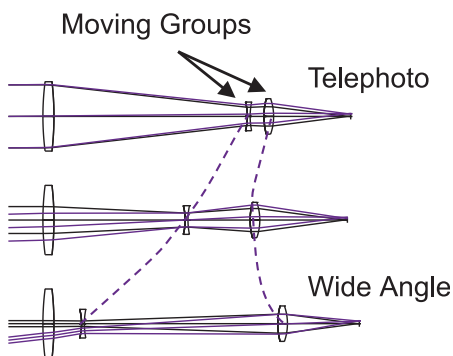
Zoom Lens Design and Optimization

Zoom lenses can be classified according to their distribution of group power (e.g., $+ - + +$). A simple **mechanically compensated** zoom lens has two moving groups: the **variator** changes magnification, and the **compensator** moves to maintain image plane focus. The motion of the compensator group is typically nonlinear; the zoom groups are linked with a mechanical **cam**. The two zoom groups form a **zoom kernel** that relays a fixed object distance to a fixed image plane. In photographic objectives, a front **focusing** group is used to provide a fixed image position for the zoom kernel and to accommodate changes to the object distance. More-complex zooms add a fourth (fixed) **prime group** to the rear of the zoom kernel and/or more moving groups.

Zoom Cam

Optically compensated zooms move linked groups using a linear motion but do not have a perfectly fixed image plane. Many discrete zooms use a form of two-position optically compensated designs.

Zoom lens optimization tips:

- Start with a thin lens layout (e.g., first-order zoom lens layouts can be designed with **Gaussian brackets**) or a predefined starting point; global optimization is not as useful for finding zoom lens starting points as it is for fixed focal length lenses.
- Independently achromatize each zoom group over multiple conjugates.
- Optimize over at least 3–5 zoom positions simultaneously and then verify the lens performance between the optimized zoom positions.
- Check that moving groups do not collide with each other or fixed groups during their zoom motion.
- More than two moving groups may be needed to maintain $f/\#$, aberration control, or telecentricity across zoom, or for thermal compensation. Adding more moving groups also helps reduce the package size for zoom lenses with large zoom ratios.

Techniques for Improving an Optical Design

Many techniques exist either to improve an existing lens with inadequate performance or to extend an existing design over new requirements without changing the basic design form. The most common scenario arises when a designer must "push" a **legacy design** to increased aperture and/or larger field (for better resolution/throughput). Such requests often come with a challenging expectation of no significant increase in the physical size of the design. Options available to a designer include:

- **Split** a lens
- **Compound** a lens
- **Asperize** a surface
- Use a **diffractive** surface
- Raise (or lower) index
- Raise (or lower) *V*
- Use an abnormal partial dispersion material

Each of the above techniques seeks to reduce the aberration contribution of a surface/component. A **Seidel** or **Pegel diagram** shows the surface-by-surface aberration contribution and is a key tool in identifying elements or surfaces to target for improvement. It can also be very useful for comparing the sensitivity of different design solutions.
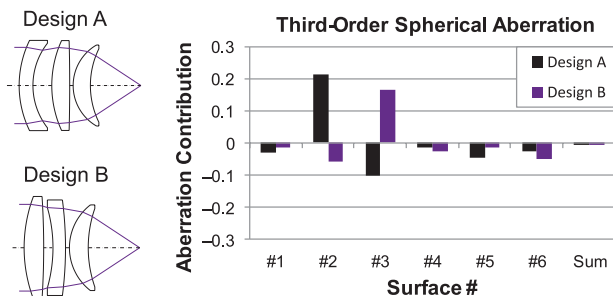
Increasing the aperture or field of an existing design can lead to **ray failures** during lens evaluation and optimization. Examples include **total internal reflection (TIR)** or **ray intercept failures** (where rays physically miss a surface). The solution is to slowly "walk" up to the desired aperture or field in small increments, optimizing the lens at each step.

Angle of Incidence and Aplanatic Surfaces
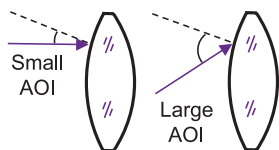
Ray **angles of incidence (AOIs)** are measured relative to the surface normal at the ray intercept. A large AOI indicates significant departure from paraxial refraction, where $\theta \neq \sin\theta$. Ray AOI is therefore an important indicator of both the aberration contributions (aberration coefficients depend on the chief and/or marginal ray AOI) and the tolerance sensitivity of a surface. During the improvement process, elements with large ray angles are good candidates for modification (e.g., splitting into multiple elements). Examining a lens layout and assessing the ray angle distribution gives a quick insight into the performance, manufacturability, and improvement potential for a design. Several "design for tolerance" optimization metrics have been proposed that minimize ray AOIs throughout a design.

Three conditions exist where a single surface can have zero third-order spherical aberration:

- The object or the image is concentric with the surface so that rays do not refract (zero AOI);
- The marginal ray height is zero at the surface (useful for field lenses); or
- The surface is **aplanatic** where the marginal ray AOI (i) equals the negative of the angle that the ray makes with the optical axis after refraction (u').

The final aplanatic condition is a powerful one, as it allows optical power and large AOIs without adding spherical aberration (or coma and astigmatism).
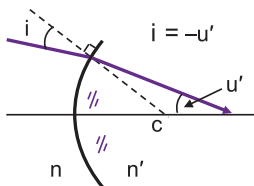
An **aplanatic lens** exploits two of the three conditions for zero spherical at a surface: the first surface is aplanatic, and the second surface is concentric with the image plane. Microscope objectives and interferometer reference spheres use aplanatic lenses to increase their speed without introducing significant aberrations.

Splitting and Compounding
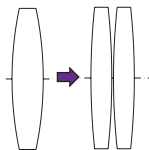
Splitting a lens element into two (or more) elements of the same material while maintaining the total power is a highly successful design technique for improving an optical system. It

- Adds degrees of freedom for optimization.
- Reduces the ray AOIs by reducing the individual surface powers and is therefore very effective at improving spherical, astigmatism, coma, and distortion.
- Reduces spherical aberration when splitting positive elements in high-NA lenses.
- Reduces astigmatism, distortion, and coma when splitting negative elements in wide-angle lenses.
- Does *not* improve Petzval or primary color since the total power contribution does not change.

The exact method to split a lens (e.g., literally cutting the lens in half or using two equi-convex elements) is design dependent. If the change in element shape is not significant enough, the lens will revert to its original local minimum (after splitting, the result will be two lens elements with a narrow crack-like airspace between them whose outer shape mimics the shape of the original element). However, if the change is too large, ray failures can occur during optimization.

Compounding a lens element into a doublet or triplet (while holding total power) is another powerful design improvement technique. It
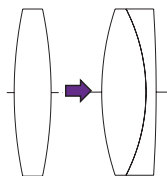
- Uses two (or more) materials to simulate a nonexistent but desirable material type (e.g., a material with a high index, large V, or unusual partial dispersion).
- Primarily reduces axial color, lateral color, and Petzval, although the extra degrees of freedom can also be used to correct other aberrations.

Diffraction-Limited Performance

During the design improvement process, performance targets are useful for guiding the process and ultimately knowing when to stop optimizing. Many optical systems use diffraction to set the performance limit, and the lens requirements may state "diffraction-limited by design." However, different definitions of diffraction-limited exist, so a designer should clarify the intent of the statement.

Strictly speaking, a **diffraction-limited lens** is limited solely by diffraction and refers to a zero-wave RMS OPD; however, these criteria are rarely the intent of a "diffraction-limited" specification. Common approximations for diffraction-limited performance include:

- a peak-to-valley OPD $< \lambda/4$ (Rayleigh criterion),

- an RMS OPD $< 0.070\lambda$ (Maréchal criterion), or

- a Strehl ratio > 0.8.

Reversing a lens can help a stagnated design during optimization. Swapping two surfaces of an element in the lens design program often requires adjusting the surrounding thicknesses so that the principal planes and first-order properties remain constant and so that ray failures do not occur. Changes in ray intercept heights and AOIs will alter the aberration distribution.
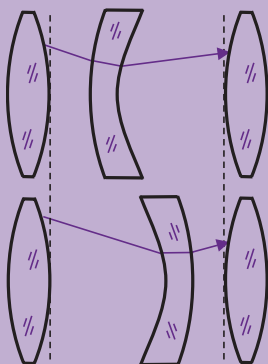
Thin Lens Layout

For a lens whose thickness is much smaller than its focal length, the marginal ray height is approximately the same at both surfaces. The ray transfer contribution through the element can be ignored and the lens treated as a **thin lens** with zero thickness. A **thin lens layout** simplifies a complex optical system by making each element (or lens group) a thin lens with optical power. This layout highlights the balance of positive and negative power between lens elements and/or lens groups and can be a vital aid to understanding the performance limitations of an optical design, especially if the design is limited by Petzval or color. A thin lens layout is also a useful tool for starting an optical design (essential for zoom lenses) or joining independent lens modules together while matching their pupils in size and location.

For a thin lens, the third-order aberration equations can
$$\beta = \frac{c_1 + c_2}{c_1 - c_2} \quad C = \frac{u_a + u_a'}{u_a - u_a'}$$
be simplified and written as a function of a **lens bending parameter** or **shape factor** β that describes what the lens looks like (defined by the two curvatures), a **conjugate factor** C that describes the magnification at which the lens is used (defined by the input and output marginal ray angles), and its index of refraction. The Kingslake **G-sums** are similar **thin lens aberration expressions**, where each aberration coefficient includes a "G" coefficient that depends solely on index of refraction.

A *y-ybar diagram* is a plot of the marginal ray height y versus the chief ray height $\bar{y}$, where each point on the plot represents a lens element (or surface) in the system. The plot quickly identifies the locations of objects

Three-Element (A, B, C)
Erecting Afocal Telescope

and images ($y = 0$) and stops and pupils ($\bar{y} = 0$) and their sizes, and is an alternative to a thin lens layout.
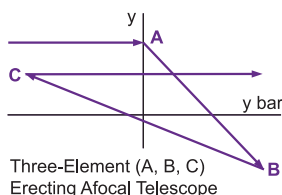
Lens Bending

A key variable when optimizing a design is lens radius. For a specific index of refraction and thickness, there are an infinite number of combinations of radii R_1 and R_2 that will yield a lens with a given focal length. The result is that for a fixed focal length, a lens may take on any number of different shapes or "**lens bendings**," and the aberrations of the lens (primarily spherical aberration and coma) will change as the shape is changed.
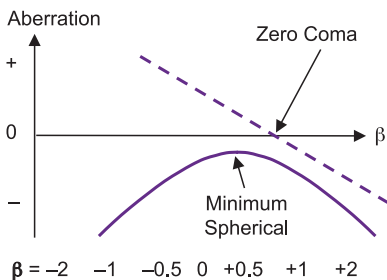

For C = –1 (object at infinity):

Minimum Spherical
$$\frac{c_2}{c_1} = \frac{2n^2 - n - 4}{n(2n+1)}$$

Zero Coma
$$\frac{c_2}{c_1} = \frac{(n^2 - n - 1)}{n^2}$$

The thin lens aberration equations can be used to show that for a single thin lens with an object at infinity (conjugate factor $C = -1$) and the stop at the lens, a bending parameter β exists *either* for minimum (but nonzero) spherical *or* for zero coma that depends only on the index of refraction of the lens, whereas astigmatism, Petzval, and distortion do not depend on the shape of the lens.

The shape factor for a single thin lens (object at infinity) for minimum spherical aberration changes from near convex-plano to meniscus as the index of refraction is increased from 1.5 to 4.0. This behavior helps explain why very high-index infrared lenses "look" different from visible lenses.

f/2
100-mm EFL

N-BK7
$\lambda = 0.55$ μm
$n = 1.52$
$\beta = 0.62$

Germanium
$\lambda = 4$ μm
$n = 4.0$
$\beta = 5.3$
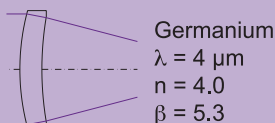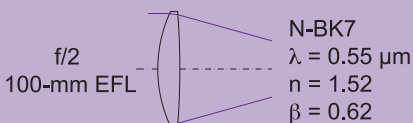
Material Selection

Material selection is one of the most critical choices in lens design. Changing a lens material during the optimization of a design can significantly alter its performance. In general, raising the **index of refraction** of all elements is an effective way of reducing the lens curvatures for given lens powers, minimizing ray AOIs and therefore aberrations at each surface. Primary chromatic aberrations can be corrected by selecting materials with the appropriate **dispersion (V)**. In a thin

lens **achromatic doublet**, the *V difference* between the two materials determines the individual element powers. Increasing the positive element *V* and decreasing the negative element *V* will reduce the component

$$\Phi = \phi_1 + \phi_2$$
$$\phi_1 = \Phi \frac{V_1}{V_1 - V_2}$$
$$\phi_2 = -\Phi \frac{V_2}{V_1 - V_2}$$

powers, the ray AOIs, and the corresponding aberrations. **"Old" achromats** use "lead-line" glasses with a positive low-index crown and a negative high-index flint (e.g., BK7/SF2). **"New" achromats** use a high-index positive crown and a lower-index negative flint (e.g., LaKN17/F2). A new achromat has ~4× less Petzval than an old achromat, but the small *V* separation due to material limitations restricts it to about half the speed of an old achromat. New achromats are useful in designs driven by field performance (e.g., wide-angle lenses).

The secondary spectrum of a thin lens achromatic doublet is proportional to the $\Delta P / \Delta V$ ratio of the two materials, where *P* is the **partial dispersion** of a material. Ideally, a pair of glasses with the same *P* but significant difference in *V* is needed to correct secondary spectrum. Unfortunately, for most glass doublet combinations, the $\Delta P / \Delta V$ ratio is nearly a constant. Only a few materials (e.g., CaF_2) exist with **anomalous partial dispersion** that helps reduce secondary spectrum. A triplet achromat can correct secondary spectrum by using two of the materials in the triplet to "synthesize" a material that can be matched in partial dispersion with the third material.

Controlling the Petzval Sum

In order to have an optical design with a reasonably flat field, its **Petzval sum** must be small. The Petzval sum is directly proportional

$$\sum_j \frac{\phi_j}{n_j}$$

to the ratio of each element's power to its index of refraction. The reciprocal of the Petzval sum is the **Petzval radius**, which equals the radius of the ideal image surface. The ratio of the Petzval radius to the focal length of the lens is the **Petzval ratio**.

Designing a lens to have zero Petzval is difficult, as positive focal length lenses want to have primarily positive-powered elements. A key strategy to control the Petzval sum is to separate positive power from negative power:

$$\Phi = \phi_1 + \phi_2 - \frac{t}{n}\phi_1\phi_2.$$

For example, if the radii are equal in a single lens (a meniscus lens), the surface powers are equal but opposite sign, resulting in a lens with zero Petzval and a power that is directly proportional to the thickness of the lens. In more-complex lenses, where positive lens groups are separated in space from negative lens groups.

In most positive-power refracting lenses, the Petzval surface will be curved toward the lens (inward-curving field). In **catadioptric designs**, reflective surfaces can be used to balance the Petzval from positive lenses because a positive mirror has an index of refraction of –1, yielding an overcorrected (backward-curving) Petzval contribution equal to its optical power.

Petzval curvature is perhaps the hardest aberration to correct in an optical system. Once the first-order power balance has been established, the index of refraction is the only variable available to control the Petzval sum. In positive elements, a higher index reduces the inward-curving Petzval; however, it is not so clearly defined in negative elements because a lower index increases the overcorrecting Petzval, but a higher index reduces the surface curvature, reducing other aberrations.

Stop Shift and Stop Symmetry

Finding the best position for the stop is an important part of the design process. A **stop shift** (moving the stop location and scaling its size to maintain the $f/\#$) forces the chief ray through different parts of a lens. Aberrations that depend on the chief ray (coma, astigmatism, distortion, and lateral color) will change, whereas spherical, Petzval, and axial color are unchanged with stop shift. For a thin lens with nonzero spherical aberration, there is always a stop position that exactly eliminates coma (it also minimizes the tangential field curvature). This position is called the **natural stop position**.

> If the best stop location is not obvious, a **floating stop** can be utilized during optimization. The first surface in the lens is designated as a dummy "stop" surface with a variable negative distance. This surface is really the entrance pupil location and allows the aperture stop surface to "float" to its best location. Before finalizing the design, the aperture stop surface designation should be moved to the correct location to account for the effects of any pupil aberrations.
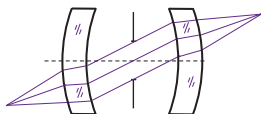
 Distortion, coma, and lateral color can be eliminated in lenses with **stop symmetry**. These aberrations all have an odd-order dependence on the chief ray height (or angle), so lenses on one side of the stop have an aberration contribution equal and opposite to lenses on the other side. In practice, even lenses with stop symmetry still have some residual aberrations because the object/image conjugates are not usually symmetric. Stop symmetry is a key strategy for reducing distortion (similar to Petzval, distortion is a notoriously difficult aberration to correct).

> Unlike film, digital sensors cannot accept large ray AOIs and typically have chief-ray-angle (CRA) matching requirements. This trait can restrict the positioning of the aperture stop in the system.

Telecentricity

A **telecentric** lens is a lens with either the entrance pupil (**object-space telecentric**) or exit pupil (**image-space telecentric**) at infinity. The degree to which the chief ray is parallel to the optical axis at either the object or image plane is called **telecentricity**. The simplest way to make

a lens telecentric is to place the aperture stop at a focal point of the lens.
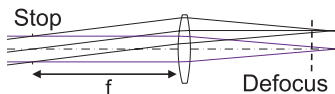
In a design code, the object space is easily defined to be telecentric; however, the chief ray must be constrained to go through the center of the stop during optimization. A telecentric image space can be attained by targeting the inverse of the exit pupil distance to zero in the merit function or with a CRA solve or CRA constraint (usually at several field points).

The primary benefit of telecentricity is that the system magnification is insensitive to slight changes in image distance (focus). The incident cone angle is also constant over the entire field, improving **relative illumination**. This is important for microlithography, microscopes, machine vision systems, and other types of metrology lenses. However, the lens elements must be at least as large as the image, making telecentric lenses larger than their nontelecentric counterparts.

A **4f optical system** is a **doubly telecentric** finite-conjugate relay (also known as an **afocal relay telescope**) with an accessible pupil plane for **Fourier filtering**. Each of the two lenses has a focal length of *f* so that the total distance from object to image is 4*f*.
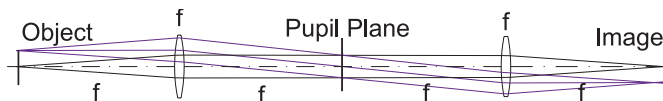
Vignetting

By definition, the aperture stop limits the size of the imaging bundle for an on-axis object point. **Vignetting** occurs when another surface (optical or mechanical) in the system clips image-forming rays for off-axis field points. The amount of vignetting is measured as a percentage of the width of the pupil. Vignetting reduces the amount of light reaching the image plane for off-axis field points. Some lenses (e.g., lithographic) cannot tolerate any light loss or drop in **relative illumination** across the field, while others (e.g., camera lenses) allow as much as 75% vignetting.

Unvignetted

Vignetted

20% Vignetted   50% Vignetted
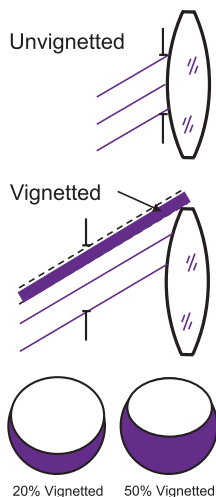
Optical design codes simulate vignetting by approximating the vignetted pupil shape as an ellipse. **Vignetting factors** relate the scaling and offset of the elliptical pupil to the unvignetted pupil. (Note that some programs use a percentage and an offset, and others define both an upper and a lower percentage.) **Pupil aberrations** may lead to confusing "negative" vignetting factors (even in the absence of physical vignetting) because the paraxial entrance pupil must be overfilled to evenly fill the aperture stop surface.

The minimum clear aperture for zero vignetting at any surface is the sum of the absolute value of its marginal and chief ray heights ($|y_a| + |y_b|$). If allowed, vignetting is a powerful technique to improve the image quality of a design because the pupil rim rays at the edge of the field typically cause the largest spot flare. Vignetting also reduces element diameters that are too large to meet packaging and/or weight constraints. Any clipping surface should be far enough away from the stop so as to not clip the axial beam during optimization. Stop shifts can help increase (or decrease) the amount of vignetting.

Pupil Aberrations

The pupils in an optical design are not necessarily perfectly imaged to each other. To derive third-order equations for **pupil aberrations**, the entrance and exit pupil planes are treated as

Field Aberrations		Pupil Aberrations
Object	⟷	Entrance Pupil
Image	⟷	Exit Pupil
Marginal Ray	⟷	Chief Ray
Aperture Stop	⟷	Field Stop

conjugate planes (similar to the object/image surfaces), and the marginal and chief rays in the system are interchanged in the third-order field aberration equations. Although pupil aberrations do not directly impact image quality, they can strongly influence other critical lens properties. For example, large pupil aberrations make it hard to **pupil match** two optical systems (especially objectives and eyepieces) together.

Nagler Eyepiece: f/5, 15-deg HFOV

Pupil spherical is the most commonly encountered pupil aberration. In wide-FOV eyepieces, pupil spherical causes the exit beams from different field points to not cross at a well-defined axial location for an observer's eye. The result is that eyepieces with large pupil spherical may have a shadowing or **"kidney-bean" effect** in the apparent brightness of off-axis fields. Pupil spherical can also cause ray tracing **ray failures** if there is a large shift between the paraxial entrance pupil location (used for the initial chief ray aiming) and the full-field pupil location. Design codes can calculate pupil aberration ray fans and/or help determine if **ray aiming** (which forces explicit x/y coordinates for the chief ray on a given surface) is required to continue with the design analysis.

Distortion is less sensitive to changes in object position if the pupil spherical in a lens is small—an important trait in photoreducers or magnifiers that require well-corrected distortion over variable conjugates.

Aspheres: Design

By definition, **aspheres** are "nonspherical" surfaces. Standard aspheres are rotationally symmetric, but off-axis conic sections and **freeform surfaces** are also referred to as aspheres. Many

Dual Gull Wing Asphere
w/ Inflection Points

qualitative terms exist that help describe an aspheric shape (e.g., mild, weak, steep, or **gull wing**). In optical design, aspheres are used to increase performance or reduce the size and weight of a system. Near a pupil, aspheres can be used to correct all orders of spherical aberration. Near an image plane (e.g., on a field lens), they can be used to correct astigmatism or distortion, or achieve telecentricity. Early applications for aspheres included illuminators, infrared systems, and single-wavelength, high-NA, small-field systems like DVD objectives or laser diode collimators.

> When optimizing with aspheres, a designer should significantly increase the pupil ray density and, if the asphere is near an image plane, add more field points.

The **sag** of a surface is the displacement (along z) of the surface from its vertex as a function of lateral position (x, y, or radial coordinate r). The traditional **aspheric sag equation** combines a **conic** expression with an even-order polynomial, yielding a rotationally symmetric surface with no vertex discontinuity:

$$sag = z = \frac{cr^2}{1 + \sqrt{1 - (\kappa + 1)c^2 r^2}} + dr^4 + er^6 + fr^8 + gr^{10} + \cdots$$

Both the **conic constant** κ and the r^4 coefficient will affect third-order spherical, whereas the higher-order coefficients are useful for correcting higher orders of spherical aberration (in high-NA systems). However, unneeded high-order terms can create oscillatory surfaces and optimization convergence problems. Alternatively, a **Q-type (Forbes) polynomial** aspheric expansion is an orthogonal basis set where the coefficients are linked to aspheric departure and slope, making it more practical for tolerancing and manufacturing.

Aspheres: Fabrication

Fabrication advances have made high-quality aspheres practical for almost any design. The **aspheric departure** is a measure of the surface deviation from a purely spherical surface and is one of the most critical fabrication metrics. A lens drawing typically specifies both a **base** or **vertex radius** and a **best-fit sphere (BFS)**. The BFS minimizes the amount of material to be removed. Several different techniques exist for fabricating aspheres (each with their own pros and cons).

Technique	Material	Comments
Conventional polishing	Glass / Crystals	High-quality surfaces; long cycle time
Ion milling	Glass	High-precision removal; slow removal rates and limited departures
Deterministic polishing	Glass / Crystals	High-quality surfaces; mid-spatial "ripple" residuals
Molding	Glass / Plastic	Low-cost, high-volume replication (high NRE for molds); limited materials and lower surface quality
Diamond turning	Plastic / Crystals	Wide range of surface shapes; serial throughput; scattering/diffraction from tool marks

Maximum **slope error** (the slope is the first derivative of surface departure) is another metric that drives both the fabrication and the testing of an asphere. Slightly oversizing the aspheric clear aperture during the design will keep the large slope errors that tend to occur at the edge of the part outside of the critical aperture; however, the total departure and part diameter will increase. **Aspheric testing** can contribute significantly to the overall part cost. Standard interferometric testing can be used with small-departure surfaces, but other techniques are required for large-departure surfaces.

Test Method	Comments
Surface profilometry	1D low-resolution scan; probe contacts part (risk of damage)
Null lenses and computer-generated holograms (CGHs)	High accuracy; costly and specific to a single surface; expensive and long lead time; precision alignment required
Subaperture stitching	Useful for a wide range of aspheres; full 2D surface coverage; limited slope
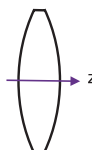
Gradient Index Materials

A **gradient index** or **GRIN** material has a spatially varying index of refraction typically described by an index polynomial. However, the index polynomial is not directly related to the manufacturing parameters, leading to lengthy development cycles. **Axial gradients** have planes of constant index perpendicular to the optical axis. Their index equation is convergent; only the linear term N_{01} is needed to correct all orders of spherical aberration when combined with a spherical surface. The result is that axial gradients make fast collimating lenses (similar to aspheres).

Change of Index
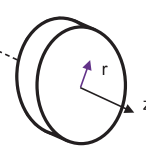
Radial gradients have cylinders of constant index centered on the optical axis. Radial gradients have the unique ability to focus light with flat surfaces. A thin radial gradient is called a **Wood lens**, and its power is proportional to its thickness times its N_{10} coefficient. Radial gradients are used in fiber couplers and 1:1 relay systems such as endoscopes, boroscopes, and desktop copiers (replacing complex lens assemblies with a smaller, less-expensive single component). A **quarter-pitch** rod focuses incoming collimated light at its back surface, whereas a **full-pitch** rod relays an upright image from its front surface to its back.

$$n(z) = N_{00} + N_{01}z + N_{02}z^2 + \dots$$

$$n(r) = N_{00} + N_{10}r^2 + N_{20}r^4 + \dots$$

GRIN aberrations are split into a **surface contribution** into the medium and a **transfer contribution** within the medium. Axial gradients have large surface and small transfer contributions, whereas radial gradients have large transfer and small surface contributions. The primary manufacturing process for GRIN materials is ion exchange. Although convenient for modeling and aberration analysis, the index polynomial is not directly related to manufacturing parameters, leading to lengthy material development.

Diffractive Optics

A **diffractive optical element (DOE)** uses wavelength-scale surface or phase structure to redirect an optical beam using diffraction rather than refraction or reflection. **Diffraction efficiency** is the fraction of light diffracted into the desired order. Low diffraction efficiencies can cause significant stray light problems. DOEs work with both refractive and reflective surfaces and, similar to conventional elements, they have a focal length, dispersion, and aberrations. Similar to aspheres, DOEs can be used for aberration correction or system size/weight reduction. DOEs can also generate optical power from flat surfaces or be used for athermalization, but they are usually more expensive than aspheres.

> Diffractive surfaces have a constant *negative V* (equal to −3.45) that can be used to correct primary color. However, their abnormal partial dispersion yields a much larger secondary spectrum than a standard doublet.

Kinoforms are smoothly varying surface-relief DOEs typically made by diamond turning a curved substrate. Kinoforms can be replicated in high volume and are common in infrared systems as hybrid diffractive achromats, where the DOE replaces the negative

Kinoform of Order m

$$h = \frac{\lambda}{m(n-1)}$$

flint element. A **binary optic** is a "stair-step" DOE with a discrete number of steps in either the surface or phase profile and is formed by multiple lithography steps (the diffraction efficiency is directly proportional to the number of steps). A **holographic optical element (HOE)** is formed by interfering two wavefronts. A **computer-generated hologram (CGH)** is a HOE formed by directly writing a phase profile on a

Binary Optic (4-step)

substrate. Both types of HOEs can represent arbitrary phase profiles. CGHs are often used as **null lenses** to test large departure and/or off-axis aspheric surfaces.
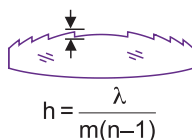
Optimization

Given an optical design starting point, modern lens-design programs use **numerical optimization** to find design solutions with improved performance by minimizing a predefined **merit function**. Also called a **cost** or **penalty function**, the merit function uses a single number to represent total lens performance; in general, smaller numbers denote better design solutions.

Designers first identify the lens **variables** (e.g., radii, thicknesses, airspaces, and material parameters) that are allowed to **vary** during optimization; all other lens parameters are kept **frozen**. Performance **defects** (or **operands**) that depend on the lens variables are then defined and used to construct the merit function. Example operands include image-quality measures (e.g., spot size) and first-order properties (focal length). **Constraints** (e.g., minimum lens thickness) can also be applied that limit the amount of change in a variable.
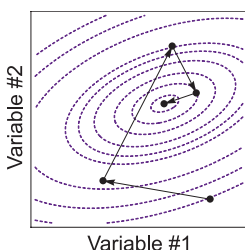
Because the operands in the merit function usually outnumber the variables in the optical design, optimization algorithms attempt to find merit function minima, not exact solutions. The merit function topology and solution complexity depend on the number of variables. A lens with only two variables has a 2D topographical solution space, whereas systems with more lens variables will have multiple minima in a multi-dimensional merit function. Commercial software programs use proprietary local and global optimization algorithms to find minima in the merit function.

Local optimizers use gradient methods to find the "nearest" minimum. The starting variable values determine which local minimum is found. **Global optimizers** are allowed to break out of local minima to find a more-comprehensive solution set.

Damped Least Squares

The most common method of **local optimization** is **damped least squares (DLS)**, which minimizes the merit function using linear regression approximations. In DLS, each variable is changed by a (small) **step size** that balances good numerical accuracy with fast convergence time. Calculating the change in each merit function operand for each variable step change produces a **derivative** or **gradient matrix** and defines a local slope that allows the optimizer to move "downhill" and find a better solution using a numerical gradient search.

An optimizer can stagnate before reaching a good design solution when the merit function changes very little with the variable changes. Manually checking the merit function for overconstrained variables, changing the operand weighting, or dramatically altering one or two variable values (e.g., radius of curvature or airspace) can often overcome stagnation. In some software programs the variable step sizes (also called **derivative increments**) and/or the **damping factor** can be modified to prevent stagnation and improve the convergence of the optimizer.

Standard least-squares algorithms assume linear dependences between targets and variables. However, in lens design, the relationship between merit function operands and lens variables is highly nonlinear. DLS also approximates the functional dependence between operands and variables as linear but its
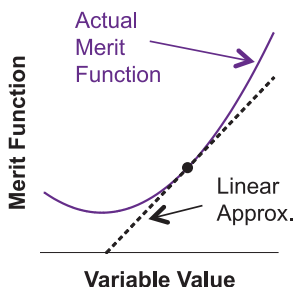
damping factor adds a weighting factor to the merit function that heavily penalizes any large changes in the variables. If the changes are small, then a nonlinear function approaches a linear approximation. Although the damping factor may slow the optimization, it is more robust for highly nonlinear targets.

Global Optimization

Local optimization uses gradient search to find the nearest merit function minimum and moves "downhill." **Global optimization** attempts to find the global minimum by allowing both uphill and downhill movement in the merit function. However, global optimization can require extensive computation time.
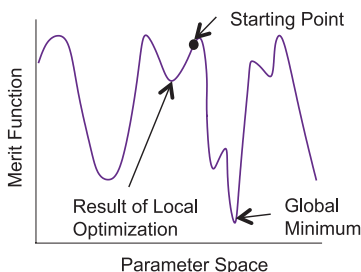
In many cases, the global merit function has a reduced set of image performance operands and more physical constraints (e.g., maximum ray AOI) than a local merit function in order to save time but keep the design solutions realistic. The result is that global optimizers do not generally find the absolute best merit function minima. Global search designs should be followed by a more-comprehensive DLS local optimization to find the optimum design solution.

Adaptive simulated annealing algorithms mimic the physical process of cooling excited atoms and molecules, where the system has a nonzero probability of reaching a higher energy state (moving uphill). **Genetic algorithms** encode lens variables in a numeric sequence and selectively choose "genes" from best-performing "parents" to create a new system.

Global optimizers are most often used as **global search** algorithms to rapidly scan parameter space for multiple solutions. Global searches require only a minimal starting design—a series of plane parallel plates with a properly defined merit function can yield good results. However, there is no guarantee that global optimization will ever find the absolute global minimum for a given problem.

Merit Function Construction

In optimization, a **merit function** is a single number that captures all aspects of desired lens performance. The merit function ϕ is constructed by taking the RMS of all identified operands, where

$$\phi = \sum_{i=1}^{m} w_i^2 (c_i - t_i)^2$$

m = number of operands
w_i = weighting factor for operand i
c_i = current value for operand i
t_i = target value for operand i

Squaring each operand serves to magnify the operands with the worst performance and ensure that positive and negative operand values do not offset each other in the sum. Individual operands are relatively **weighted** to emphasize their desired contribution to overall performance. The target value for most operands is zero.

Most design programs have a **default merit function** that automatically defines a set of image quality operands and includes the appropriate pupil sampling and weighting for fields and wavelengths.

Merit function operands consist of both image quality measures (e.g., RMS spot diameter) and construction parameters (e.g., focal length or magnification). Ray position errors and/or OPD values are used to calculate the image quality measures. For example, a simple image quality operand is the distance in the image plane of a single ray from the **centroid** (or "center of mass") of a large number of rays, all traced from the same object point. Construction parameters can be included in the merit function as heavily weighted operands (to ensure that they are met), but this approach can significantly slow down improvements in image quality. **Lagrange multipliers** are an alternative to heavily weighted merit function operands and place the construction parameter constraints outside of the merit function, allowing for more stable optimization; however, each Lagrange multiplier constraint requires one free variable.

Choosing Effective Variables

Certain variables are more effective at improving lens performance than others. **Radii of curvature** and **airspaces** are almost always left free to vary during optimization. Packaging or manufacturing constraints (e.g., requirements for a flat surface, equal radii, or working distance) may restrict their range of values.

In contrast, **lens thicknesses** are much less effective variables (a change in a lens thickness is often equivalent to a change in a nearby airspace). Element thicknesses are often kept frozen during a majority of the design process. However, in some cases (e.g., thick-meniscus, Petzval-correcting lenses), the thickness of a lens is an essential variable in improving image quality; thicknesses may also need to be released to meet edge thickness constraints. If allowed to vary, optimizers tend to make lenses very thick (or very thin), slightly improving Petzval with only a small improvement in the merit function.

> A lens element whose thickness is chosen to maintain a diameter-to-thickness or **aspect ratio** between 10:1 and 5:1 is often easier and less costly to manufacture.

"Glass" choice is a critical variable for color correction as both the index of refraction and dispersion are dependent on the material choice. Glass differs from most variables as it is a **discrete** rather than a **continuous variable**. **Model** or **fictitious glasses** allow continuous index and Abbe number for optimization, but a real material must eventually be substituted from a **glass map**.

Aspheric coefficients are important variables for aberration correction in modern designs. If possible, aspheric coefficient variables should be enabled during optimization one at a time (starting with the lowest orders), one surface at a time. It is not necessary in most designs to vary the conic constant simultaneously with the fourth-order polynomial coefficient; rather, vary the conic on strong surfaces and the fourth-order term on flatter surfaces.

Solves and Pickups

Solves define a surface curvature or thickness to be a function of a given system requirement. As the lens changes, a solve modifies the selected parameter to maintain the system requirement. Solves are typically used in the initial setup of a lens, but some lens requirements such as focal length can be constrained using solves rather than being defined as merit function operands. This method decreases the number of independent variables and corresponding constraints during optimization, reducing computing time. Examples include:

- A **marginal ray angle solve** or *f*/# **solve** on the curvature of the last surface can be used to maintain the system focal length for a fixed entrance pupil size.
- A **marginal ray height solve** on the thickness of the last surface can be used to locate the image plane at paraxial focus. If an internal image exists, this solve can also be used to find and hold its location.
- A **chief ray angle solve** on the curvature of the last surface can force image-space telecentricity.
- **Aplanatic solves** force the aplanatic condition on a surface by changing its curvature.
- **Thickness solves** can be used to control lens edge thicknesses or the total distance between separated surfaces (useful in zoom or multi-configuration designs).
- A **magnification solve** adjusts the object (or image) distance to maintain a system magnification.

Many marginal-ray and chief-ray solves require ray information based on the entrance pupil and cannot be placed on surfaces before the aperture stop.

A **pickup** is a special type of solve that "picks up" or uses data from another surface. It maintains a set relationship between two lens parameters as one of them is varied. Pickups are very useful for optimizing double-pass or symmetrical systems (e.g., to allow a lens to change its radius yet maintain symmetry, the first radius is "picked up" by the second radius as its negative value).

Defining Field Points

A lens should be optimized at multiple **field points**. The number and location of the field points heavily influence the optimization process and the final design result. To ensure that the lens performance does not oscillate between field points, a final design should be evaluated at more field points than were used for optimization, especially if aspheric surfaces are used on field lenses.

For a rotationally symmetric lens, the field is traditionally divided into circles with equal areas. The normalized radius of each circle is the **normalized field coordinate** H. In lenses with moderate field (<25-deg FFOV) and no aspheres, three fields ($H = 0, 0.7,$ and 1.0) are generally sufficient for optimization. For systems with a rectangular sensor, it is common to circumscribe a circle around the sensor whose diameter equals the diagonal of the sensor, and this is used to define the field. Asymmetric systems (e.g., TMAs) require fully defined rectangular fields in all four quadrants because no rotational symmetry exists. In some cases, lateral or reflection symmetry reduces the number of fields required for optimization.
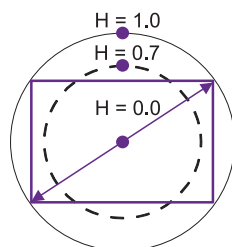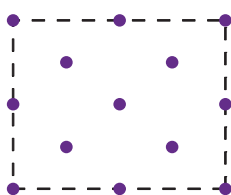
Common field definitions are **object height**, **object angle**, and **image height**. In complex systems, using image height fields often requires **ray aiming** of **reference rays** to avoid ray trace failures.

Fields can also be given **field weights** to balance or emphasize on- or off-axis performance for a given application. While projection and lithography lenses require uniform image quality over the entire field, visual and photographic systems often allow some performance loss toward the edge of the field.

Pupil Sampling
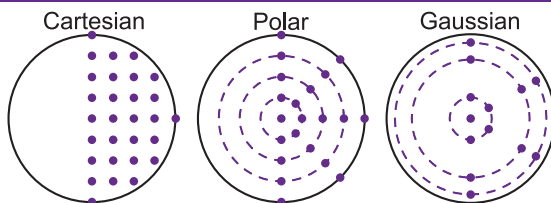
Pupil sampling defines the number and the distribution of the rays traced through the pupil and is critical to both optimization and analysis. Tracing fewer rays leads to faster optimization but also to reduced accuracy. Obscurations and vignetting will influence the required pupil sampling and the algorithm used to generate rays. Pupil sampling methods may weight rays differently according to their pupil coordinate. The pupil sampling ray density should be increased for systems with aspheric surfaces or large aberrations.

Common pupil sampling methods include:

- **Rectangular grid sampling** assumes no symmetry and often requires a high-density grid 10×10 to obtain accurate sampling. Rays are generally evenly weighted.

- **Polar** or **hexapolar sampling** uses periodically spaced radial and azimuthal rays, and generally requires fewer rays than rectangular sampling. Rays may be weighted by **ring** (concentric circle) or **arm** (radial line).

- **Gaussian quadrature (GQ) sampling** uses a very small number of **skew rays** at very specific pupil coordinates and weightings. GQ sampling returns a mathematically exact integral of the pupil with fewer rays and provides higher sampling near the edge. GQ is the fastest sampling for the majority of cases.

GQ sampling works best with circular or elliptical pupils and may not give accurate results with highly obscured or very noncircular pupils (unless specifically modified in the design code for this special case).

Tolerancing

Tolerancing is a statistical process that determines the allowable change or **tolerance** in each of the lens **construction parameters**. *All* lenses must be toleranced before being built. The toleranced design needs to simultaneously satisfy the performance requirements, minimize manufacturing costs, and minimize alignment complexity. These goals tend to be contradictory, so a large part of tolerancing is finding the best compromise among them.

Tolerances include both optical print values (e.g., radius of curvature, center thickness, index variation, wedge) and mechanical print and assembly values (e.g., tilt, decenter, axial spacing, subassembly alignment). **Compensators** are parameters that can be adjusted during the lens build (e.g., focus, airspace changes, active element centering) to recover performance losses caused by other tolerances. Tolerances are often separated into **symmetric** (compensated by focus and space adjust) and **asymmetric tolerances** (compensated with image tilt and push-arounds). Tolerances that cannot be easily modeled (e.g., material inhomogeneity or complex surface figure errors) form part of an unallocated **performance budget**, requiring extra **design margin** for the unknown variations. Many different tolerance methods exist, but they generally follow the same basic process:

1. Choose initial tolerance values for all parameters.
2. Define the performance metrics (e.g., MTF, RMS spot diameter, boresight error) *and* the requirements.
3. Run a sensitivity analysis to determine the impact on performance from each tolerance, and identify sensitive and cost-driving tolerances.
4. Define compensators and their allowable ranges.
5. Run appropriate statistical analyses (e.g., **Monte Carlo analysis**) and evaluate the expected as-built performance and manufacturing yield.
6. Adjust tolerances and compensators until cost and performance goals are met or a redesign is needed.

Design Margin and Performance Budgets

Design margin is the difference between the required as-built performance and the nominal design performance. The design margin should be large enough to allow for all manufacturing errors and

environmental changes. However, the best design for manufacturing is not always the one with the largest design margin. A lens with slightly worse nominal performance is usually less sensitive to manufacturing errors, resulting in a much higher **yield** and lower cost of production. Designers determine the most manufacturable design by comparing both tolerances and predicted as-built performance of each design solution.
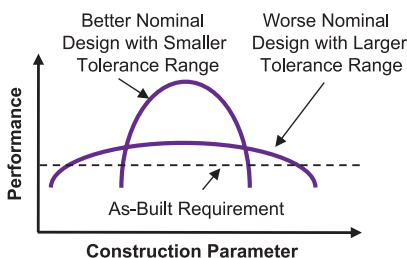
The design margin for a single top-level requirement can be allocated to several categories in a **performance budget**. The allocations in each category are then either summed or root-sum-squared (RSS) to a top-level number. In addition to tolerancing, the budget values are used for design targets, mechanical structure design (including thermal management and resonance), and other environmental requirements. There may be more
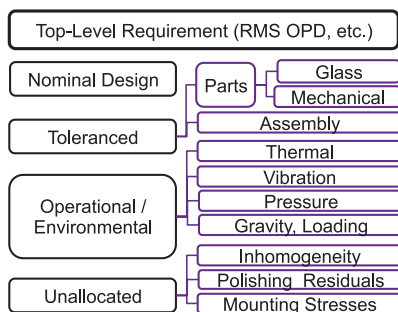
than one performance budget necessary (e.g., RMS OPD for image quality and pointing accuracy for boresight). Complicated lenses with many subassemblies may also require a performance budget where each subassembly is broken out separately.

Optical Prints

An **optical drawing** or **optical print** consists of all of the design and tolerance information necessary for the fabrication of the part, including:

- Material index of refraction, dispersion, and quality
- Radii and surface figure (power and irregularity)
- Center thickness (CT) and edge thickness
- Special surface parameters (e.g., aspheres)
- Surface finish (roughness) and quality (scratch-dig)
- Outer diameter (OD), surface sag, and edge beveling
- Wedge or total indicator runout (TIR)
- Clear aperture (CA) per surface
- Antireflection (AR) or other coating requirements

Talk with lens fabricators early in the design process to discuss tolerance requirements, as drawing formats and achievable tolerances vary from vendor to vendor. It is also important to make sure that the vendor's measurement methods support your specifications.

Optical prints typically isolate singlets or cemented doublets or triplets. Higher-level assembly tolerances (e.g., group or barrel axial displacement, and group tilt and decenter tolerances) are listed on **assembly prints** or **assembly drawings**.

ISO 10110 is the international standard for optical drawings, although many organizations develop their own customized drawing formats.

Left surface:		Material specification:		Right surface:	
R 50.0000 +0.03 (cx)		GLASS TYPE: NBK7		R 108.0000 +0.03 (cx)	
⌀ₑ 22.00		Nₑ 1.5187 +0.001		⌀ₑ 22.00	
3/3		Vₑ 63.96 +0.800 %		3/3	
4/		0/		4/-	
5/		1/		5/	
6/-		2/		6/-	
Coating:				Coating:	

Rev level	Dr.	Date			
Next assy	Chk		Title: Singlet		
Revisions	Eng		Size A	Drawing number Singlet	
Ltr descr dt	QA	Scale 4:1	Units=mm	Ind. acc. ISO 10110	
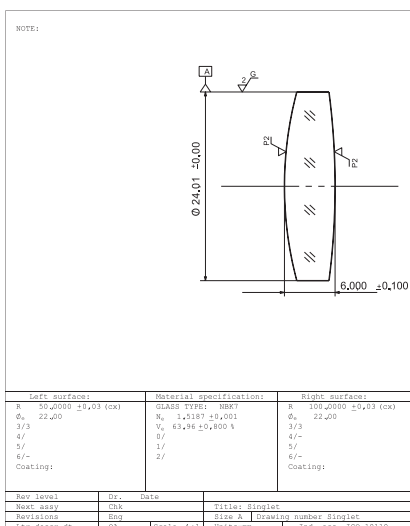
Radius of Curvature Tolerances

Radius of curvature (ROC) tolerances refer to the change in vertex curvature of a surface and are specified differently depending on how the lens is to be measured. Some designs tolerance ROC directly, assuming a measurement of ROC with either a **spherometer** (contact measurement) or a **distance-measuring interferometer** (noncontact measurement). Other designs use **test plate fringes** to indirectly tolerance ROC. In this case, it is common to list two tolerances for ROC, a radius tolerance and a test plate fringe tolerance, which can appear confusing. The radius value tells the fabricator to use a test plate within a certain radius range, and the number of fringes gives the allowable part deviation from the test plate.

> To compare the sensitivity of different radii in a design, do not apply a "uniform" radius tolerance in length units across the lens; instead, use a uniform tolerance of curvature ($c = 1/R$) or percentage of radius (%R).

Test plates are precision sur-faces used during polishing to monitor both ROC and surface irregularity. They are placed in contact with the lens and viewed under monochromatic light (typ-ically a 546.1-nm mercury lamp), forming fringes (or "**Newton's rings**").

Sag Error   Fringes

Surface Under Test   Test Plate

The ROC can be calculated from the part diame-ter and either the number of fringes (flat test plate) or the curvature of the fringes (spherical test plate). Test plate fringes represent a double-pass measurement, where one fringe corresponds to a half-wave of surface deviation. De-signs are usually **test-plate fit** to a specific supplier's cat-alog of test plates to avoid large test plate tooling costs. Each radius in the design is individually altered to match the closest catalog test plate value, and the entire lens is re-optimized after each change.
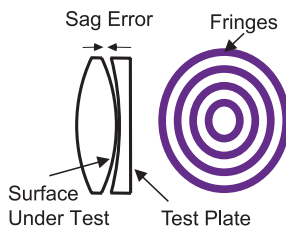
Surface Irregularity

Surface irregularity refers to any measured deviation of an optical surface from its intended shape. During a standard polishing process, irregularity is monitored with a test plate by assessing the number of fringes and/or their irregularity. For more-accurate, noncontact measurements, a surface-measuring interferometer is used to analyze the surface irregularity.

> Thin elements can "spring" when removed from a polishing block, resulting in two surfaces with large but approximately equal and opposite cylinder error. If the clear aperture on each surface is about the same, the effect of such cylinder error on lens performance nearly cancels when used in transmission.

The wide range of possible "shapes" for surface irregularity makes it difficult to model and tolerance deterministically, and it is often simply specified in fringes or waves of **peak-to-valley (P–V)** error. **Power** and **cylinder** (astigmatism) are common low-order errors produced by standard polishing processes, whereas higher-order errors include an "edge roll" or "edge rip." Irregularity specification and measurement are sensitive to the exact CA called out. Allowing extra room between the CA and the outside part diameter makes it easier to achieve high-quality surfaces.

Simple tolerance models assume that the irregularity error is pure cylinder. Compensation schemes may include rotating or **clocking** an element. More-complex tolerance models use a combination of **Zernike polynomials**, either toleranced individually or as an aggregate RMS value of all coefficients up to a certain order. Irregularity from sources such as mounting stress must also be considered, especially for mirrors. Mechanically induced surface deformation often has three-way symmetry ("**3-point**" or **trefoil**) from kinematic mount designs.

Center Thickness and Wedge Tolerances

A **center thickness (CT)** tolerance specifies the allowable error in the vertex thickness of an element. Fabricators typically leave CT on the thick side of the tolerance range to allow for removal of cosmetic defects or rework during final polishing. The sensitivity of a design to a CT error is often similar in magnitude to the sensitivity of the design to an airspace error on either side of the element.

CT
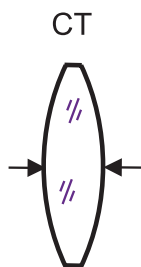
"Thin" lenses with diameter-to-CT ratios greater than about 15:1 typically require special handling during fabrication, increasing their manufacturing cost.

A **wedge** tolerance in a spherical element indicates the amount of allowable tilt between the optical axis of its two surfaces (defined by connecting their centers of curvature) and a mechanical axis [usually the **outside diameter (OD)** of the element]. Wedge is measured on a spin-

Wedge

c_2 c_1

ning chuck, either by optical methods such as angular beam deviation or by mechanical methods such as **total indicator runout (TIR)** or **edge thickness difference (ETD)**. The fabrication process of **centering** or **edging** a lens reduces element wedge and sets the final part diameter. Aspherical elements require special treatment (both for tolerance modeling and fabrication) because they do not have a clearly defined optical axis between their two surfaces.
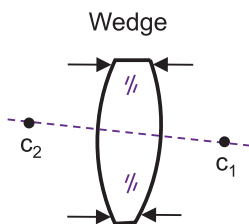
Meniscus lenses with nearly concentric radii are extremely sensitive to wedge errors. They can also be difficult to center because it is hard to remove the wedge without significantly reducing the part diameter and encroaching on the CA.

Material and Cosmetic Tolerances

The various parameters required to specify optical materials must also be toleranced in a lens design. The two most common tolerances are **index of refraction** *n* and **dispersion** *V*. A variation of $n \pm 0.001$ and $V \pm 0.8\%$ is standard for most commercial glasses. Tighter tolerances are available at higher cost. High-performance designs must also consider index **inhomogeneity** and **birefringence**. Inhomogeneity is the P–V variation in bulk index throughout the lens volume and is typically less than $\pm 5 \times 10^{-6}$ for commercial glass. Birefringence is defined as the maximum OPD between polarization states over a standard path length (typically less than 20 nm/cm for commercial glass). The complexity of modeling the inhomogeneity and birefringence errors often push them into part of the unallocated performance budget, but such errors can be toleranced effectively as surface or wavefront irregularity.

Many lenses are overspecified for cosmetic errors, leading to reduced yield and additional cost. If the lens surface is not near an image or visible to the user, demanding cosmetic specifications may not be required.

Cosmetic defects include **scratches**, **digs**, **pits**, and **staining**. Such defects generally have little influence on imaging performance, with a few exceptions:

- Defects in lenses sensitive to stray light will scatter light onto the focal plane and may add background noise. The amount of light intercepted by the defect is proportional to the ratio of the defect area to the beam area and is generally scattered over 4π sr.
- Defects in lenses for high-power laser applications may act as damage centers or stress intensifiers.
- Defects near an image surface or subtending a large fraction of the beam footprint on a given surface may appear in the image.

Cosmetic defects and bulk material defects (e.g., **bubbles**, **inclusions**, and **striae**) are hard to tolerance accurately and are often part of the unallocated performance budget.

Lens Assembly Methods

To properly tolerance a lens assembly, it is important to anticipate how the lens elements will be mounted and aligned. In general, as the required lens centration and axial spacing tolerances get tighter, the mechanical complexity, lens assembly time, and cost increase. Adhesives and flexures can be used to reduce mounting-induced surface deformation. Common assembly methods include (from low to high precision):

- **"Drop-in" assembly**: Elements are "dropped" into a lens barrel, separated by spacers, and held in place with a mechanical retainer. The assembly tolerances are driven by the achievable fit between the element OD and the barrel inner diameter (ID). These assemblies work best with low-cost targets in high-volume applications where ease of assembly is paramount.

- **Shim-centered assembly**: Elements are centered radially in a lens barrel using shims and then bonded in place. This method is used for lenses that require both tight centration tolerances and large radial gaps to allow for thermal expansion.

- **"Poker-chip" assembly**: Each element is accurately aligned to the reference surface of an individual lens cell and bonded in place. Cell and spacers are inserted as a stack into a precision-machined lens barrel.

- **Bolt-together assembly**: Each element is accurately aligned to the reference surface of a precision-machined lens cell and then held in place on a seat with either adhesive and/or flexures. The lens cells are precision-aligned to each other either optically or mechanically and then bolted together.

"Drop-In"   "Shim-Centered"

"Poker Chip"   "Bolt Together"
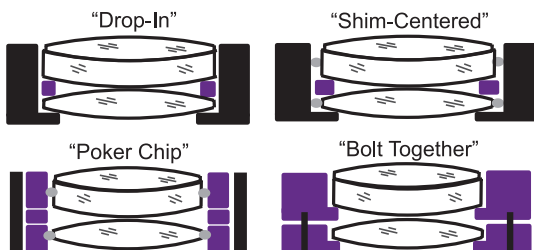
Assembly Tolerances

Assembly tolerances, such as decenter, tilt, roll, and axial spacing, represent positioning errors of optical surfaces in a lens system. Assembly tolerances may be applied to lens surfaces, lens elements, and/or element subassemblies. Accurate tolerancing requires knowledge of the optomechanical design and mechanical part tolerances, and a well-defined **assembly plan**.
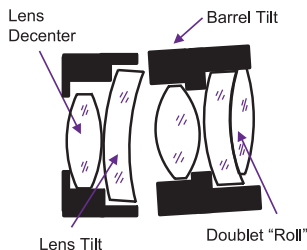
> Multiple decenter errors can be added together and toleranced as a single motion. However, multiple tilt errors with their own individual pivot points must be toleranced as separate motions.

Most assembly tolerances are not derived from a single part error but must be calculated from multiple part errors using a **tolerance stack-up**. For example, in a simple mount, a single **airspace tolerance** can depend on five separate part tolerances: two lens surface or bevel sags, a spacer thickness and diameter, and a lens CT. The impact of a CT error on an airspace also depends on how the element is mounted. Elements separated by spacers shift all subsequent surfaces by their CT errors, thus maintaining the airspaces between elements, whereas the CT error in an element mounted on an independent mechanical shoulder in a lens barrel also changes the neighboring air space value but may not affect the position of other elements.

Tilt and **decenter** tolerances break rotational symmetry introducing **boresight error** and **field tilt** as well as axial coma, axial astigmatism, and axial lateral color. To properly apply a tilt tolerance, it is important to include the pivot point for the tilt motion. For cemented elements, the tilt and decenter of one element with respect to the other is specified as **roll** about the cemented interface.

Compensators

Manufacturing errors that have a similar effect on performance can often be corrected by modifying a single lens parameter. **Compensators** are lens variables that simulate adjustments made to the lens during assembly and test to improve as-built performance. The most sensitive tolerance parameters are usually selected as compensators. Compensators loosen tolerances and reduce part cost but add time (and cost) to assembly and test. Without compensators, most tolerances would be unreasonably tight. For each compensator, the tolerance process must define the **adjustment range** and its required resolution and accuracy. The optomechanical design typically constrains the compensator **range of motion**.

> Compensators should be *limited* in number, *independent* (two compensators should not affect/correct the same aberration), and *practical* (e.g., glass dispersion is not a realistic compensator).

- **Focus** and **image plane tilt** are almost universally used as image plane compensators to loosen tolerances.
- A custom spacer may be designated as a post-build **airspace compensator** whenever focus is insufficient to control the spherical aberration introduced by **symmetrical tolerance** errors.
- **Asymmetrical tolerance** errors typically cause axial coma. This can be corrected during test by decentering a single lens element known as a **push-around**.
- In complex lenses, several airspaces can be optimized prior to assembly using the as-built measured data (radii, CT, airspaces, and index) in a **space adjust**.
- Tight index-of-refraction or dispersion requirements may require a **melt-recompensation** process, in which the indices of refraction of the actual pieces of material to be used in the assembly are measured. The design is then re-optimized (usually only the airspaces are varied) using the measured glass material data. This method can help loosen material tolerances.

Probability Distributions

The tolerance extremes for each lens variable limit only the maximum and minimum values. The **probability distribution function (PDF)** defines the likelihood of obtaining an as-built value between the extremes. The tolerance PDF depends on the fabrication and/or alignment process (which may even be specific to a given supplier) and should be defined for each tolerance in a **Monte Carlo tolerance** analysis.

> CT tolerances often have skewed PDFs because polishers will leave lenses on the thicker side of nominal to allow for possible rework due to cosmetic errors.

- **Uniform distributions** are constant over the tolerance range with equal probability between the extremes.

Uniform

- **Normal** or **Gaussian distributions** have a defined mean and standard deviation, where the mean is the nominal design value and is the most likely value to occur.

Normal/Gaussian

- Because normal distributions allow infinite range, a **truncated Gaussian** is often used, where the probability is defined to be zero beyond the tolerance extremes. This distribution represents a supplier yield where parts out of specification are not used.

Skewed

- **Skewed Gaussian distributions** have a nonzero higher-order moment, where the mean or nominal value is not the most likely value.

Spiked or Binary

- **Spiked** or **parabolic distributions** are heavily weighted toward the toler-

Parabolic

ance extremes, where the nominal design value is the least likely. They can also simulate binary tolerances.

- Most design codes permit the use of **custom distributions** to incorporate measured statistical data.
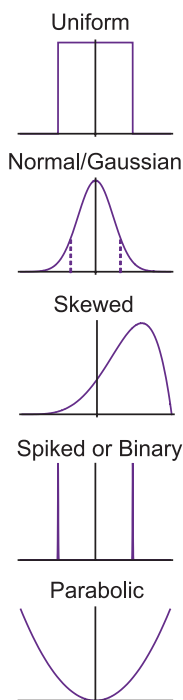
Sensitivity Analysis

A **sensitivity analysis** determines the sensitivity of a lens to manufacturing errors by calculating the change in each performance metric (e.g., RMS wavefront or MTF) for both a plus and a minus change (equal to the tolerance value) in each tolerance parameter. Performance changes may not be symmetric about the nominal design value and often vary with field point. The analysis is initially performed **uncompensated** (although focus is regularly included in the initial run). If needed, a second **compensated** run (where the compensator motion is optimized to improve performance) helps gauge the effectiveness of the compensator and identifies tolerances that drive large compensator motions. During the compensated run, constraints are often removed from the merit function (e.g., focal length) that can interfere with performance improvement.
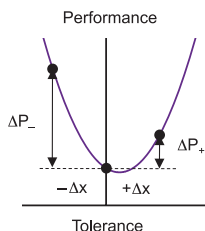
An **inverse sensitivity analysis** determines the tolerance value that produces a given change in performance, with the goal of having each tolerance contribute equally to the performance degradation of the system.

It is often useful to apply the same tolerance value to all surfaces (or elements) for common tolerance parameters (e.g., CT ±0.025 mm or index ±0.001) and then group the results into tables and/or charts. A **sensitivity table** or **chart** lets you identify at a glance the tolerances that will drive the performance loss and highlight possible compensators. The tables and charts are also very useful in comparing the sensitivities of different design solutions to possible manufacturing errors.
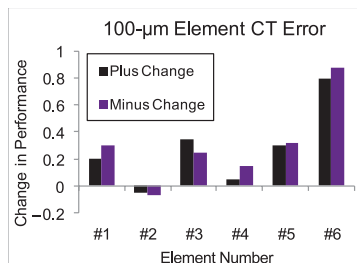
Performance Prediction

A basic sensitivity analysis considers the effects on system performance for each tolerance individually. In order to predict the image quality of an as-built system, it is necessary to simulate the effect on performance of all of the tolerances simultaneously. Most design software packages compute the as-built **performance prediction** using a **root-sum-square (RSS)** calculation, displaying sensitivity analysis results in tabular or graphical form. This approach works well if the performance losses from all of the tolerances are of similar form.

The RSS of the performance change from each tolerance gives an estimate of overall predicted performance; however, it generally overestimates the performance loss when compared to a Monte Carlo analysis with the appropriate PDFs and tolerance interactions (e.g., a lens radius may have a different sensitivity if the radius on the other side of the lens also changes).

In some cases, part tolerances are divided into **symmetric** (surface-centered) or **asymmetric** (non-surface-centered) **tolerances** for separate tolerance runs:

Symmetric
- Curvature (radius)
- Center thickness
- Airspace
- Index of refraction
- Dispersion

Asymmetric
- Surface tilt/decenter
- Lens tilt/decenter
- Group tilt/decenter
- Surface irregularity
- Index inhomogeneity

Wavefront differential or **derivative tolerancing** generates a multivariable differential tolerance **sensitivity matrix** containing first-order (linear), second-order (quadratic), and mixed partial derivatives or **cross terms**. It requires a one-time analysis of the nominal system and can be hundreds of times faster than a full Monte Carlo analysis.

Monte Carlo Analysis

A **Monte Carlo analysis** is an alternative tolerancing approach to predict as-built performance. It simulates a manufacturing environment by generating a series of lenses or **trials**. Each trial run randomly perturbs all tolerances within its defined extremes according to its PDFs, and then evaluates the image performance criteria. By considering all applicable tolerances simultaneously and exactly (no approximations are made), a highly accurate simulation of expected performance is possible. If compensators are needed, each trial can be reoptimized with only the compensators as variables. If allowed, relaxing exact constraints (such as focal length or working distance) in the merit function during compensation can significantly loosen the symmetric tolerance values.

> A rule of thumb for a Monte Carlo analysis is that the number of trials should exceed the square of the number of tolerances for good statistical validity.

The result of a Monte Carlo analysis is a **probability distribution table** and/or **cumulative distribution function (CDF)** or **yield curve** showing the percentage of lenses that meet or exceed a given performance criterion. The number of random trials evaluated determines the maximum statistical confidence level. Because tolerancing is statistical, a designer needs to decide in advance what is an acceptable **manufacturing yield** (e.g., 98% pass and 2% fail). Individual trials can be saved for forensic analysis or to investigate poorly performing outliers.
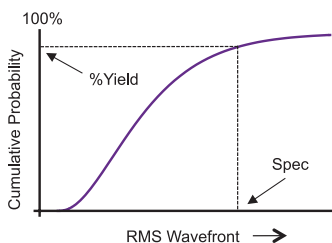
Monte Carlo analyses generally require significant computation time and can only give statistical results, which may not be valid for one-off builds or very small production runs.

Environmental Analysis

A complete tolerance analysis must include the performance impact on a lens due to changes in its environment. **Environmental requirements** such as temperature or pressure can consist of multiple specification ranges. The **operating range** dictates the total extent over which the lens must meet all performance requirements. A **survival range** gives the extent over which the lens need not meet performance but cannot break or become nonfunctional. The sensitivity of the lens to environmental perturbations must be assessed and allocated as part of the performance budget using an **environmental analysis**.

In addition to ambient **temperature** changes, heat sources such as electronics, motors, light absorption, and solar radiation can all induce thermal effects in lens systems. Two key material properties are needed for an accurate thermal model: The **coefficient of thermal expansion (CTE)** gives the percentage change in linear dimension per unit temperature change, and the glass **thermo-optic coefficient** gives the absolute change in refractive index per unit temperature change. Simple tolerance models specify the temperature change as a thermal soak or steady-state effect; however, more-complex models can include both spatial and temporal temperature gradients. **Pressure** or **altitude** changes primarily affect the refractive index of the gas material surrounding the lens elements (typically air) but can also alter the airspaces between lenses or change the shape of optical surfaces in sealed assemblies.

Most design codes provide the ability to perturb the lens temperature and/or pressure and evaluate performance changes as long as the appropriate material data is entered. However, complex environmental influences need **finite element analysis (FEA)** to help predict changes in performance [e.g., **vibrations** can disrupt component alignment and need FEA to model the vibration effects on **line-of-sight (LOS)** stability and **image jitter**].

Athermalization

As the temperature of an optical system changes, the lens elements and housing physically expand/contract. In addition, the glass refractive index will also change. The primary result of both changes is a focus (or image plane) shift, although aberrations and pointing errors (from lateral component shifts) are also possible. An **athermalized** optomechanical design minimizes performance variation with temperature. For example, focus shift can be **actively athermalized** by measuring the temperature and then physically moving either the lens or the sensor. This approach is not the ideal solution in many situations, and **passive athermalization** (no motor-driven component movements) is employed with the use of dissimilar materials (e.g., optical materials with different thermo-optic coefficients or metal mounting materials with different CTEs). Most infrared materials have large thermo-optic coefficients, making athermalization mandatory for large temperature ranges.
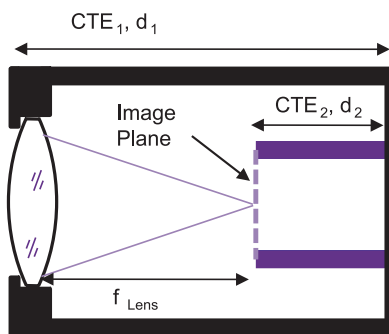
$$\frac{df_{Lens}}{dT} = CTE_1 d_1 - CTE_2 d_2$$

A reflective system can be passively athermalized against isotropic thermal soak conditions by making the mirrors and the mounting structure out of the same material. The system then scales in dimension as the temperature changes but maintains overall focus and imaging performance.

Stray Light Analysis

Stray light is any unwanted light that reduces performance by striking the image plane of an optical system. Primary sources of stray light include reflections from uncoated or poorly coated surfaces, surface or bulk scattering, thermal emission, bright out-of-field objects, glints from mechanical structures, diffracted light from edges, and unwanted diffraction orders from gratings.
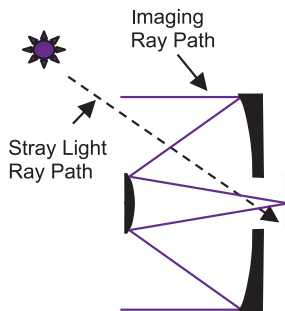
Stray light can be divided into two main categories: **ghost images** and **veiling glare**. Ghost images have structure (often with sharp edges) and create local image-quality disruptions. Veiling glare produces a diffuse background haze on the image plane, resulting primarily in a loss of overall image contrast.

> Tracing rays backwards from the detector (or "putting your eye at the detector") is a highly efficient technique for isolating potential stray light paths and is much more effective than tracing from a source to a detector.

A **stray light analysis** determines both the amount of stray light at the detector and the different paths that it took to get there. If the stray light is large enough to cause a problem, the analysis should also identify any design changes needed to reduce the stray light. Some key metrics used to quantify stray light include **off-axis rejection (OAR)**, **point-source transmittance (PST)**, **point-source normalized irradiance transmittance (PSNIT)**, and **Narcissus-induced temperature difference (NITD)** in infrared systems. The accuracy of a stray light analysis is limited by the availability of source and scatter data, the completeness of the computer model, and the available computer time.

Stray Light Reduction

Identifying potential stray light paths during the design phase is essential to avoiding costly errors in hardware. A **critical object** is any surface that can be seen by the detector, including surfaces that are imaged through lens elements. An **illuminated object** receives power directly from the source and depends on the source location. The overlap between the list of critical and illuminated objects highlight key stray light paths. An object on both lists is called a **first-order stray light path**, which is a major contributor to stray light and should be addressed. **Second-order stray light paths** exist anywhere that light can propagate (transmit, reflect, scatter, or diffract) from an illuminated object to a critical object.

Field stops reduce the number of illuminated objects in a design because objects behind the field stop cannot be seen by the source—only objects in front of the field stop can be illuminated. Similarly, placing the **aperture stop** as close as possible to the detector reduces the number of critical objects in a design since objects in front of the aperture stop cannot be seen by the detector. This approach may require that the aperture stop be relayed from the front of the lens system to the rear because the farther the aperture stop is from a lens, the larger the front elements become in diameter. **Lyot** or **glare stops** are apertures placed in conjugate pupil planes; such stops might need to duplicate any aperture stop obscurations from mounting features for extreme stray light control.
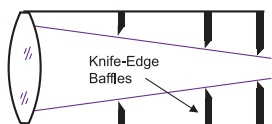
Knife-Edge Baffles

Identified stray light paths can be eliminated by blocking light with **baffles**, **vanes**, and **knife edges**, or by relocating an object so that it is no longer seen by the detector or illuminated by a source. Mechanical surfaces and edges of optical elements can be treated to absorb or scatter light by blackening or roughening. **Rifling** or **threading** barrels and mechanical surfaces is a common and economical way of reducing stray light.

Antireflection (AR) Coatings

Optical surfaces require **antireflection (AR) coatings** to reduce stray light and improve transmission. For an uncoated surface, the amount of **Fresnel reflection** loss depends on the angle of incidence of the light and the index of the lens material. For light normally incident on a surface in air, the reflection loss is ~4% for a visible glass with an index of 1.5. In the infrared spectrum, the loss per uncoated surface can be as much as 36% due to the high index of infrared materials.

Large changes in ray AOIs across a lens surface can lead to highly nonuniform pupil transmission and pupil polarization in high-NA systems. Steep ray angles of incidence on an AR-coated surface will also **blue-shift** the spectral region of a coating, especially coatings with sharp band edges.

The ideal single-layer AR-coating material for a glass-to-air interface has a quarter-wave of optical thickness and an index of refraction equal to the square root of the lens material index.
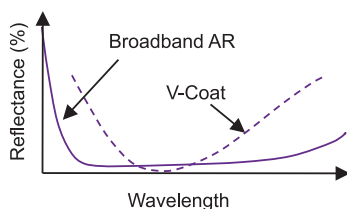
This condition is difficult to achieve in practice. The closest material choice for the visible spectrum is a single quarter-wave layer of MgF_2 (index 1.38), which will reduce the surface reflection from ~4% to ~1% at a single wavelength. More-complex coatings are made of alternating layers of high and low index, each with a thickness equal to a quarter-wave optical thickness. Two- and three-layer **V-coats** can eliminate the reflection at a single wavelength, whereas more-complex **broadband AR (BBAR)** coatings reduce the reflection loss to <0.5% over an extended wavelength band. A variety of coating design optimization software packages exist (very similar to lens design). Most lens design codes will model (but not optimize) thin film stacks for accurate transmission calculations.

Ghost Analysis

Ghosts are a form of stray light from unwanted reflections. Traditional ghosts are double-image artifacts formed by two lens surfaces (**two-bounce ghosts**) with imperfect coatings. However, ghosts can also result from reflections from object, filter, detector, and mechanical surfaces. In interferometer optics, single-reflection ghosts create extra "reference" beams that complicate fringe analysis. In systems with high-power lasers, ghosts can produce an internal focus inside a lens material or on a lens surface that may actually damage the lens. In thermal infrared lenses, **Narcissus** reflections from cooled detector features cause image artifacts. A **pupil ghost** is a special ghost image that echoes the shape of the aperture stop and is typically seen in camera lenses when a bright source (such as the sun) is inside the object field.

> Windows, beamsplitters, concentric mensicus lenses, and optical surfaces with near-normal incident rays are the primary sources of in-focus ghost images.

Most of the potential **ghost images** in a lens system are highly aberrated and out of focus. If the ghost image is significantly defocused from the image plane, it produces a general haze on the detector, reducing image contrast. A

Out-of-Focus Ghost

Aberrated Ghost
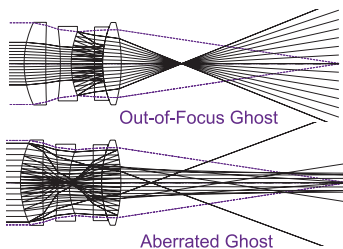
ghost analysis is primarily a graphical technique whereby sequential ray trace routines generate graphical ghost paths that analyze every potential two-bounce reflectance in the optical system and then report the "image quality" of each ghost in terms of spot size, defocus, magnification, and $f/\#$. The relative intensity of a ghost image can be estimated from its number of reflections; however, **nonsequential ray tracing** is required for a quantitative radiometric ghost analysis.

Cold Stop and Narcissus
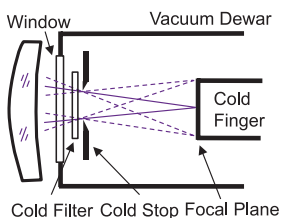
Many infrared systems require a cyrogenically cooled detector that sits in a thermally insulated vacuum **dewar**, reducing background noise and increasing sensitivity. Because the system components also emit the same wavelengths of thermal radiation as those trying to be imaged by the detector, a **cold shield** is needed to minimize the number of "warm" surfaces (outside of the object scene) that the detector can see. For optimum stray light control, the lens exit pupil should match the detector cold shield in size and location. Imaging the lens stop onto the cold shield forms a **cold stop**, and if imaged perfectly, the system is said to have 100% **cold shield efficiency**.

Surfaces inside the dewar (e.g., the focal plane and the cold stop) are at significantly different temperatures than the scene imagery, and their reflections from an optical surface will form "cold" Narcissus ghost images. Such images are either diffuse background noise or sharp image artifacts varying in field or time. Retroreflection due to a small marginal ray angle-of-incidence at a surface or image reflection due to a small marginal ray height are primary sources of Narcissus. The **YNI product** combines these into a paraxial quantity that can be used to optimize the lens or locate baffles. Staring detectors allow a **nonuniformity correction (NUC)** to calibrate out static reflections, but any lens motion or environmental changes after NUC will induce Narcissus. The **NITD** is the sum of reflected energy from all optical surfaces.
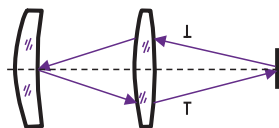
Techniques for minimizing **Narcissus** include anti-Narcissus coatings, baffles, tilting flat windows and filters, and designing with Narcissus ray constraints.

Nonsequential Ray Tracing

Nonsequential ray tracing (NRT) is an extremely powerful analysis technique for evaluating stray light. Unlike standard **sequential ray tracing** in which rays intercept surfaces only once and in a given order, NRT allows a ray to "see" all of the surfaces in a system simultaneously and the ray interacts with the first surface it meets. It is possible for a ray to intersect a single surface multiple times. NRT is also widely used to design and optimize **illumination** and other **nonimaging systems**.

For accurate results, many orders of magnitude more rays need to be traced with NRT than sequential ray traces. As a result, stray light analysis can be time consuming and computationally expensive.

A stray light analysis with NRT requires three components: **sources** that generate rays, objects that interact with rays, and **receivers** that collect the rays and quantify the resulting irradiance or intensity distribu-

tions. Key features that can be enabled depending on computation limitations are **ray scattering** and **ray splitting**, which split energy between a **parent ray** and multiple **child rays**. The sum of the transmittance, reflectance, scattering, and absorption on a surface should equal unity. Ray databases that include all ray-surface data can be saved for subsequent filtering and analysis.
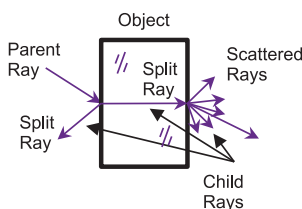
NRT models are typically much more complex than sequential optical surface design models and can include:

- multiple sources with rays propagating in any direction;
- mechanical features such as lens mounts, barrels, apertures, baffles, and vanes;
- **absorption** and **scattering** from bulk materials; and
- absorption and scattering from mechanical and optical surfaces (ground, polished, coated, or painted).

Scattering and BSDF

Optical and mechanical surfaces have a finite **surface roughness** that can induce light **scattering**.

Specular   Diffuse   Real Surface

The direction and amount of scattered light depend on the incoming ray wavelength, polarization, and angle of incidence. Because the scattered light is a potential source of stray light, a complete stray light analysis requires accurate scatter models for each surface. Ideal scatter models assume that the scatter is either all concentrated in the **specular direction** (reflected angle governed entirely by Snell's law) or is perfectly **Lambertian** (diffuse surface with the same reflected radiance over all viewing angles). Real surface scatter usually lies somewhere between these two extremes, often with most of the light scattered "around" the specular reflection direction.

The **bidirectional scattering distribution function (BSDF)** indicates how specular or diffuse a surface is as a function of both the 2D incident angle (θ_i, ϕ_i) and the 2D scattering angle (θ_o, ϕ_o):

$$BSDF(\theta_i, \phi_i; \theta_o, \phi_o) = \frac{L}{E} \left(sr^{-1} \right).$$

The BSDF is defined as exiting radiance divided by incident irradiance and is centered around the specular angle defined by $\theta_i = \theta_r$. A Lambertian surface has a constant BSDF for all angles of incidence. The reflective and transmissive versions of the BSDF are the **BRDF** and **BTDF**, respectively. A **scatterometer** is used to measure angle-resolved BSDF, BRDF, and BTDF. **Isotropic** surfaces, such as pitch-polished lenses, scatter light equally in all azimuthal directions, simplifying the BSDF function. **Anisotropic** surfaces, such as diffraction gratings, diamond-turned mirrors, or brushed-metal surfaces, have inherently azimuthal directionality. **Total integrated scatter (TIS)** is the integral of the BSDF and is a common scalar metric for total surface scattering without any angular resolution information.

Photographic Lenses: Fundamentals

A **photographic objective** (or **camera lens**) is a lens assembly that forms a real image of a real object. **Photographic lenses** were historically designed to image on film, but modern **digital cameras** use **CCDs** or **CMOS detectors**. **Camera objectives** are typically optimized for distant objects, incorporating **floating elements** to refocus the image onto a fixed plane as the user changes the object distance.

> The expressions **fast**, **slow**, and **speed of a lens** originate from photography, where a camera lens with a smaller $f/\#$ results in a wider aperture and more light collected, allowing a user to set a faster shutter speed for the same exposure time.

A wide variety of applications for camera lenses exist in consumer (e.g., point-and-shoot cameras, mobile phones, camcorders), commercial (professional movie cameras), industrial (quality control), and scientific markets. Most modern photographic objectives have zoom capability (to change magnification), and use aspheres and plastic materials for both size and cost reduction. Cameras typically operate in the visible, but they can be designed for any wavelength band where a detector is available.

Camera lenses are classified by focal length and **image format** size (e.g., 100-mm telephoto for 35-mm format). The focal length for a given image format determines the angular FOV. A "normal" lens has a focal length approximately equal to the image diagonal and a FOV of about 50 deg. A **long-focus lens** is a narrow-FOV lens with a long focal length; the most common type of long-focus lens is a **telephoto lens**. A **wide-angle lens** is a lens with a FOV wider than 60 deg and a focal length shorter than normal. **Fish-eye lenses** have extremely large FOVs (>180 deg). **Macro lenses** are special objectives designed for objects that are smaller than the image sensor (>1:1 **reproduction ratio**).

Photographic Lenses: Design Constraints

Most cameras need to be handheld, restricting the size and weight of a photographic lens. A **single-lens reflex (SLR)** camera lens has an additional working distance requirement. A diagonal mirror is located between the SLR lens and the image plane to allow the user to view through the lens itself (the mirror flips out of the way just before exposure). In contrast, a **viewfinder camera** has a separate viewfinder so that lens elements can be much closer to the image plane. However, the viewfinder FOV does not perfectly overlap the lens FOV, and "what you see" is not necessarily "what you get."

MTF (in cycles/mm) is used to specify and measure camera lens performance. Unlike telescope and microscope objectives, which are usually diffraction limited at full aperture, photographic objectives are aberration limited at full aperture and are rarely used "wide open." Stopping the system down decreases the spot size blur due to aberrations and increases the depth of focus of system. A speed of $\sim f/8$ balances diffraction and aberration effects in SLR photographic lenses.

Photographic lenses have relaxed requirements for **relative illumination (RI)**, the ratio of on-axis image plane irradiance to the minimum irradiance over the full image field. This permits designers to use vignetting to reduce

aberrations and increase the aperture and/or FOV. For wide-FOV designs, a 50% RI specification is not uncommon. Distortion requirements for lenses with standard FOVs range from <1% to 4%. However, in wide-FOV lenses (especially fish-eye lenses), correcting distortion so that image height is proportional to $f\tan(\theta)$ is very difficult; it becomes easier to control f-θ **distortion**, where equal object-angle increments map to equal image-height increments.
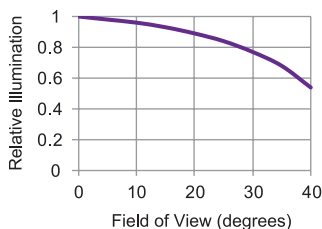
Visual Instruments and the Eye

The **human eye** is a dynamic optical system that can **accommodate** varying object distances by changing the shape and refractive power of its lens. The refractive power is typically measured in **diopters**, where 1 diopter = 1/meter. The eye forms its image on the **retina**, a curved surface containing a series of rods and cones that serve as detecting units. The resolution of the eye varies over its FOV, with the best resolution found in the central **foveal region** where the cone density is a maximum. The **fovea** subtends about a 1-deg FOV with roughly 1-arcmin angular resolution. The resolution drops outside this region (e.g., it is ~5 arcmin for a 30-deg off-axis field point).

There is a practical limit on the maximum useful magnification for visual instruments. At some point, the image becomes so large that the details in the image exceed the resolution of the eye. This is referred to as **empty magnification**. In applications where resolved detail is important (e.g., microscopy), this can lead to artifacts in the viewed image. However, in applications where the boresight of the image outline is more important (e.g., riflescopes), some amount of empty magnification can be beneficial for relieving eye strain. As a rule, to avoid empty magnification, the visual instrument magnification should not exceed ~11× the aperture diameter (in inches) in telescopes and ~225× the NA in microscopes.

Visual instruments, such as telescopes, microscopes, riflescopes, endoscopes, periscopes, and binoculars, form images that can be viewed with the human eye. The image is typically magnified and placed at a convenient location for viewing, ranging from 25 cm (10 in) to infinity, depending on the application. Simple visual instruments consist of an **objective** and an **eyepiece**, while more complex systems add **image inverters** and/or **relays**. Visual instruments should also have a focus adjustment to accommodate differences between individual users.

Eyepiece Fundamentals

Eyepieces are essential components of visual systems, magnifying the image created by an objective and placing it at a suitable location for viewing with the human eye. The eyepiece focal length determines its magnification. An eyepiece is usually designed "backwards," starting from the infinite conjugate (eye side) to the short conjugate (objective side). To avoid vignetting, eyepieces should be **pupil-matched** (in both size and location) to the human iris; this requires that the aperture stop be completely outside the lens so that the eye can be positioned a comfortable distance away from the eyepiece.

> High-power microscope objectives often have uncorrectable lateral color that can be offset with a **compensating eyepiece**.

The distance from the aperture stop to the first lens element (or first mechanical surface) is called the **eye relief**. Eye relief requirements vary from as little as 10 mm to simply clear the eyelashes, to 20 mm for eyeglass clearance, to as much as 5 in. for protection from high-power rifle recoil. For maximum light collection, the aperture stop diameter should be matched to the eye's iris diameter. Eyepieces are typically designed for a 3 to 4-mm diameter pupil, but this value may vary between 2 mm (for bright scenes) and 10 mm (for use in moving vehicles).

Due to the external aperture stop requirement, eyepiece designs have little symmetry, resulting in significant astigmatism, lateral color, and distortion (e.g., 8–10% distortion is not unusual). The off-axis image quality of eyepieces can be relatively poor—the eye dynamically scans the image during its normal operation, employing its low-quality peripheral vision in static situations. Simple eyepieces with positive elements have large uncorrected Petzval, whereas more-complex eyepieces use a negative field lens to flatten the field. The residual field curvature is typically evaluated in diopters of required accommodation (defocus).

Eyepiece Design Forms

Eyepiece design forms can be classified by number of elements, apparent FOV, and eye relief. Complex designs with more elements have improved performance, larger FOVs, and/or longer eye reliefs, but with added cost, size, and weight (important for handheld or head-mounted instruments). Modern eyepieces incorporate **aspheres** and **diffractive** elements to reduce size and weight. A key distinction among eyepiece design forms is the presence of an accessible image plane, which can be used to overlay a reticle onto the final image.

Lens distortion changes sign if the optical system is reversed. This fact is useful when balancing the distortion between an eyepiece and an objective if the eyepiece is designed in the opposite direction.
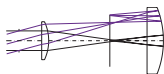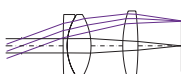
Huygens: 15-deg HFOV   Kellner: 20-deg HFOV   Plossl: 20-deg HFOV
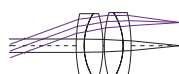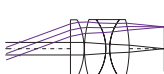
Orthoscopic: 20-deg HFOV   Erfle: 30-deg HFOV   Nagler: 40-deg HFOV
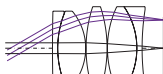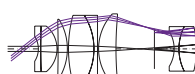
The **Huygens** eyepiece is a simple two-element design with a virtual image, while the three-element **Kellner** eyepiece has a real image plane. With two identical doublets, the **Plossl** design is a low-cost, medium-FOV eyepiece. The **orthoscopic** eyepiece (known for its low distortion) is common in astronomical and medical devices. The **Erfle** is a standard wide-FOV military eyepiece with long eye relief but short working distance. The wide-FOV **Nagler** design uses a strong negative field lens group to correct field curvature over a virtual image plane, but it suffers from strong **spherical aberration of the pupil**, leading to a "**kidney bean effect**" wherein portions of the field become alternately light and dark (vignetted) as an observer's eye position changes.

Telescopes

Objective $\quad MP = \dfrac{f_o}{f_e} \quad$ Eyepiece

f_o $\quad$ f_e $\quad$ Eye
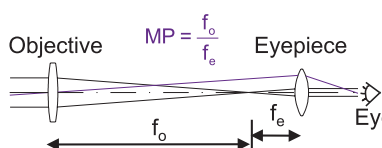
Traditional **afocal telescopes** magnify distant objects by enlarging their apparent angular extent at the eye. They consist of a long-focal-length, narrow-FOV objective and a short-focal-length, wide-FOV eyepiece. The **angular magnification** of the telescope is defined by the ratio of the objective focal length to the eyepiece focal length. Telescopes typically have interchangeable eyepieces with different focal lengths to change magnification. Modern **imaging telescopes** remove the eyepiece and use the telescope objective to image directly onto a sensor. In **terrestrial (spotting) telescopes**, an **erector** is placed between the objective and eyepiece to form an upright image.

> Afocal telescopes and beam expanders are specified by their magnification (e.g., 4×), whereas imaging telescopes are characterized by their aperture diameter (e.g., 12 in.).

Astronomical telescopes require large apertures to gather as much light as possible from distant objects. Astronomical telescopes larger than 12 in. are usually reflective. In contrast, **surveillance spotting scopes** need to be compact and portable and are usually refractive. A **riflescope** is a specialized afocal telescope requiring extended eye relief for recoil. **Binoculars** are dual co-boresighted telescopes with upright images.

Beam expanders are afocal telescopes used "in reverse" to expand a laser beam diameter and reduce its divergence. A **Keplerian telescope** (with a positively powered eyepiece) has an internal focus, allowing a possible spatial filter location. A **Galilean telescope** (with a negatively powered eyepiece) is shorter and works well for high-power lasers because there is no internal focus. A **collimator** is a reversed telescope objective.

Microscopes

A **microscope** is an instrument used to magnify and view objects that are too small for the naked eye. Applications include biological and medical research, metallurgy, and industrial inspection. A basic **compound microscope** requires an objective and an eyepiece, where the magnification is the product of the individual magnifications and is defined relative to the angle the object subtends at the eye at a distance of 250 mm. The **tube length** is the distance from the objective's rear focal point to the intermediate image.
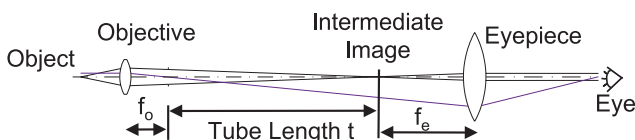
A modern microscope is designed with an **infinity-corrected objective** and a **tube lens** (placed between the objective and eyepiece). This form allows the insertion of tilted components (e.g., beamsplitters, filters, and polarizers) in the collimated space between the objective and tube lens. The tube lens can also focus the image directly onto a detector. Microscopes need to be modular (e.g., interchangeable objectives to adjust system magnification and resolution). The objective **flange-to-focus** distance is standardized at either 45 or 65 mm. The intermediate image diameter ranges from 20 to 28 mm. Standard tube lengths include 160, 180, 200, and 210 mm.

The design of the illumination system can have a significant impact on the contrast and resolution of a microscope. **Trans-illumination** uses a **condenser** to illuminate the sample (light is collected by the objective on the other side), whereas **epi-illumination** uses the objective to illuminate the sample. **Dark field illumination** (the opposite of **bright field illumination**) is a technique used to enhance contrast where the condenser has a central obscuration that is larger than the collecting aperture of the objective lens.

Microscope Objectives

Microscope objectives are small-FOV lenses that are typically diffraction limited on-axis, telecentric in object space, and unvignetted. The objective housing is imprinted with a magnification for a fixed tube lens focal length and NA (e.g., $20 \times /0.5$). Although there is no rigid relationship between magnification and NA, if the magnification is increased, the NA is increased to avoid **empty magnification**. Biological microscope objectives need to correct the spherical and chromatic aberration introduced by a thin **cover slip** or **microscope slide**.

Lister: 10× / 0.25 NA

Long WD: 20× / 0.5 NA

Amici: 20× / 0.5 NA

High Resolution: 60× / 0.9 NA
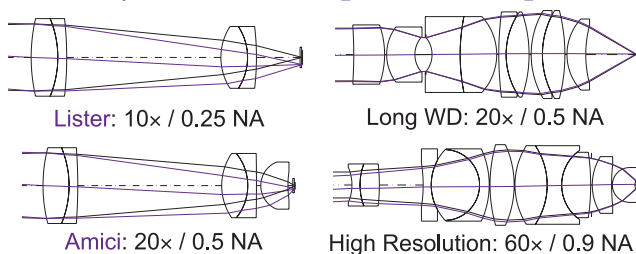
Microscope objectives are typically designed "backwards" from the long conjugate to the short conjugate. Low-power objectives are ordinary achromatic doublets dominated by field curvature. A medium-power **Lister** objective consists of two achromats in a Petzval-like arrangement. To increase the NA, **aplanatic** lenses (as in the **Amici**) are added to the object end of the lens, decreasing the working distance. Medium- to high- power objectives are limited by lateral color, which can be offset by a tube lens or eyepiece. Inverted telephoto and double-Gauss design can be turned into microscope objectives with long working distances. An **immersion objective** has an extremely high NA (>1) and is designed to interface with objects in fluid (index > 1). Objectives are classified by their field flatness and color correction capability:

- **PLAN**: corrected over a flat field (digital focal planes require flatter fields than visual instruments).
- **APO (apochromatic)**: broad-spectrum color correction.
- **PLAN–APO**: corrected for both field curvature and chromatic aberration for a high level of performance.

Relays

Finite-conjugate **relay lenses** image a real object to a real image, typically with a magnification close to one. They can be grouped into single-stage **reproduction lenses** and multi-stage unity-magnification relays, where each additional stage inverts the image.
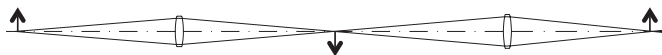
Reproduction lenses range from simple document copy/ scan lenses to complex lithography projection lenses. Their magnification varies between 1:1 and 20:1 (either reducing or enlarging) and may be adjustable. Since the applications for these lenses require a high level of distortion correction, design forms with inherent stop symmetry (e.g. the double-Gauss) are frequently used.

F-theta lenses are designed to have deliberate barrel distortion where the image height follows an $h' = f\theta$ rather than the standard $h' = f\tan\theta$ relationship. F-θ lenses are useful in scanning systems where the f-θ distortion creates a scanned spot with a constant velocity across the field, producing a uniform exposure.

Unity-magnification relays are found in **periscopes**, **endoscopes**, and **boroscopes**, where light must be guided down a long tube. Minimizing the lens diameter in such long-distance relays is critical; this requires multiple stages, each with a primary imaging lens (usually a doublet or triplet for color correction). **Field lenses** are placed near the internal image planes of the relay to converge the oblique beams so that they pass through a smaller clear aperture. Due to their proximity to an image plane, field lenses must have excellent surface quality and very few defects. The relay is typically only one part of the optical system and is coupled to a high-resolution objective on one end and an eyepiece or camera lens for viewing the image on the other end.

Index of Refraction and Dispersion

The **index of refraction** (or **index**) of a material is the ratio of the speed of light in vacuum to the speed of light inside the material (typical values range from 1.45 to 4.0). The change in index as a function of wavelength is called material **dispersion**. **Normal dispersion** occurs when index decreases with wavelength. **Anomalous dispersion** occurs near material absorption peaks where index increases with wavelength. The dispersion of a material over a given spectral band is quantified by an **Abbe number** or **V-number (V)**, which is calculated with the index at three wavelengths—the two points near the ends of the spectral

$$V = \frac{n_d - 1}{n_F - n_C}$$

Spectral Line	Wavelength
Mercury (g)	435.8 nm
Hydrogen (F)	486.1 nm
Helium (d)	587.6 nm
Hydrogen (C)	653.3 nm

band and a midpoint—which in the visible corresponds to the F, d, and C spectral lines. A high *V* corresponds to low dispersion, and a low *V* indicates high dispersion.

Custom Abbe numbers can be defined over any band of interest when designing outside of the visible spectrum (e.g., the g line is used instead of the F line when correction at shorter wavelengths is needed). The dispersion of a material can be vastly different in two different spectral bands (e.g., visible and MWIR).

Two common dispersion models are the **Sellmeier** and **Schott** index equations. Discrete measured data can be fit to one of these equations for use in optical design.

Sellmeier

$$n(\lambda) = \sqrt{1 + \frac{c_1 \lambda^2}{\lambda^2 - c_4} + \frac{c_2 \lambda^2}{\lambda^2 - c_5} + \frac{c_3 \lambda^2}{\lambda^2 - c_6}}$$

Schott

$$n(\lambda) = \sqrt{c_1 + c_2 \lambda^2 + \frac{c_3}{\lambda^2} + \frac{c_4}{\lambda^4} + \frac{c_5}{\lambda^6} + \frac{c_6}{\lambda^8}}$$

However, using dispersion equations outside of their intended wavelength region can lead to significant extrapolation errors.

Optical Materials: Glasses

Common optical materials used in the visible spectrum include **glasses**, **crystals**, and **plastics**. Of the three, glass has the most extensive historical industrial development. The primary optical properties of a glass are illustrated on a **glass map**, which plots the index of refraction of the glass versus its dispersion (typically plotted in decreasing *V*). Each dot on the glass map represents a different material. Low-dispersion **crown glasses** have a $V > 55$, and high-dispersion **flint glasses** have a $V < 55$. Glass makers further divide the glass map into regions based on proprietary chemistry. Historically, lead was used to increase the index of glass during production. The increased lead content also increased the dispersion so that the main sequence of legacy glasses lie on a "**lead line**." Glasses far from the lead line tend to be more expensive.
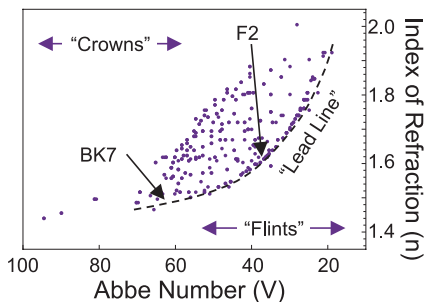
The manufacturer is the ultimate authority on an optical material. Designs with extreme material sensitivity should confirm the optical properties with an actual measurement or **melt certification**.

Major glass manufacturers now produce environmentally friendly, lead-free glasses and have **preferred glasses** that are in high production and have lower cost than standard glasses. Other important glass properties not shown on a glass map include transmission, partial dispersion, coefficient of thermal expansion, thermo-optic coefficient (dn/dT), birefringence, and glass transition temperature.

Optical Materials: Polymers/Plastics

Plastics or **optical polymers** are common in high-volume lens systems. **Injection molding** is the preferred method for fabricating low-cost lenses, but polymers can also be diamond-turned or conven-

tionally fabricated. Common materials include acrylic (polymethlymethacrylate/PMMA), polycarbonate (which can be used as a shatter-resistant material), polystyrene, and cyclic-olefin copolymer (COC/Zeonex). Plastics have material properties that are significantly different from optical glasses or UV/IR crystals.
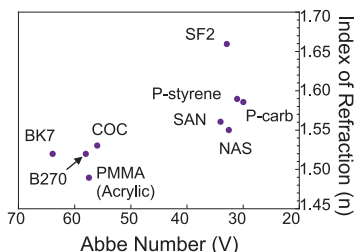
Advantages

- Inexpensive, lightweight material allows extremely high-volume and repeatable production.
- Injection molding allows precision optomechanical mounting features to be integrated with optical surfaces.
- Absorbing dopants or dyes can be included in the optical material, allowing integrated spectral filtering.
- Optical designs can easily incorporate diffractive and highly aspheric (with inflection points) surfaces.

Many glasses (e.g., Schott B270) can now be injection-molded and are cost-effective complements to polymers.

Disadvantages

- Material selection is limited compared to optical glass.
- Injection molding requires large tooling costs.
- Very low heat tolerance (<100–200 °C maximum).
- Spectral bandpass is generally only 0.4–1.6 μm.
- High thermal sensitivity: 10× larger CTE and 50× larger *dn*/*dT* than optical glass.
- Optical coatings can be difficult to apply.
- Very large stress birefringence.
- Sensitivity to abrasion and chemical exposure.

Optical Materials: Ultraviolet and Infrared

The selection of high-quality **ultraviolet (UV)** materials for UV applications (e.g., lithography, semiconductor inspection, broadband spectrometers, and fluorescence microscopy) is very limited. Most optical materials do not have good transmission below 400 nm and have increased bulk scattering. Fused silica (SiO_2) and calcium fluoride (CaF_2) are the two most-common UV materials. Other materials with high UV transmission (e.g., crystal quartz, sapphire, lithium fluoride, barium fluoride, and potassium bromide) have undesirable qualities for lens fabrication (soft/hard, birefringent, or hygroscopic). Materials with low visible dispersion are more dispersive in the UV. Many broadband deep-UV systems are reflective due to the lack of materials for color correction.

Most visible glasses are opaque (have a long-wave transmission cutoff) above 2.7 μm. Most IR materials are opaque (have a short-wave cutoff) below 1 μm.

Infrared (IR) materials are less limited in selection than UV materials. Many IR materials are **crystalline**, and a large fraction are **semiconductors**. The two main IR spectral regions are the **mid-wave (MWIR)** band from 3–5 μm and the **long-wave (LWIR)** band from 8–12 μm. Common IR materials are silicon, germanium, CaF_2, zinc selenide, and zinc sulfide. AMTIR, gallium arsenide, magnesium fluoride, and barium fluoride are also used. IR materials generally have much higher indices (as large as 4) than visible glasses. Dispersion can be very low, with *V* numbers as high as 1000 in the LWIR. The longer wavelengths reduce surface scatter and roughness requirements, making IR materials attractive for cost-effective aspheres. However, the material itself can be very brittle or soft.

IR materials have much higher **thermal expansion** and **thermo-optic coefficients** than visible glasses, and **athermalization** is required to maintain the lens performance over large temperature changes.

Snell's Law and Ray Tracing

Snell's law (of **refraction**) is the primary equation used in ray tracing. It governs the change in direction of a ray at a dielectric interface between two materials. Snell's law relates two angles in one plane of incidence; a full 3D ray

$$n\sin\theta = n'\sin\theta'$$

trace typically uses **direction cosines**, the ray's unit vector projection onto each coordinate axis. Because of the sine functions in Snell's law, ray tracing calculations through multiple surfaces become highly nonlinear for large angles. Snell's law of reflection can also be applied to a reflecting surface if the sign of the index of the incident material is reversed after the **reflection**.
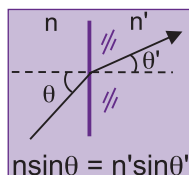
Once the ray is refracted or reflected at a surface, it must be **translated** (or **propagated**) to the next surface, changing its height (or location in 3D) without changing its angle. **Paraxial ray tracing** assumes that for small angles, $\sin(\theta)$ is approximately equal to θ; this significantly simplifies ray tracing equations [replacing the $n\sin(\theta)$ with $n\theta$ in Snell's law]. The resulting **paraxial refraction** and a **paraxial transfer equation** allow visual **brick** (or **ynu**) **diagrams** or matrix math to be set up quickly to calculate the first-order properties of an optical system.

Refract

$$n'u' = nu - y\phi$$
$$\phi = c(n' - n)$$

Transfer

$$y' = y + n'u'\left(\frac{d}{n'}\right)$$
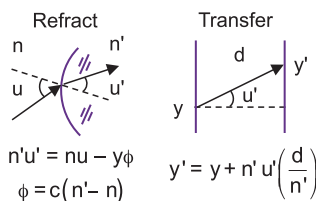
Ray failures generally occur when either a ray fails to intercept the next sequential surface or the ray undergoes total internal reflection (TIR) at a surface. Optical design codes have significant difficulty evaluating a lens system if reference rays (e.g., the chief ray) fail to trace.

Ray misses   Ray TIR
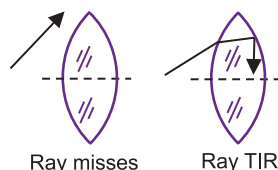
Focal Length, Power, and Magnification

Focal length is a funda-mental first-order prop-erty of a lens. For a thin lens, given the object dis-tance and focal length, the **Lens maker's equa-tion** can be used to quickly find the image distance. The focal length also determines the **mag-**

$$\frac{1}{s'} = \frac{1}{f} + \frac{1}{s} \qquad m = \frac{h'}{h} = \frac{f}{s+f}$$

$$\phi = \frac{1}{f} = (c_1 - c_2)(n-1)$$

nification (relative size) between the object and image for a given object position. The **power** of a lens is the recipro-cal of its focal length. For a **thin lens**, the power is related solely to its two curvatures and its index. For a thick lens, an **effective focal length (EFL)** is defined that equals the distance between its **focal** and **principal planes** in either object or image space.
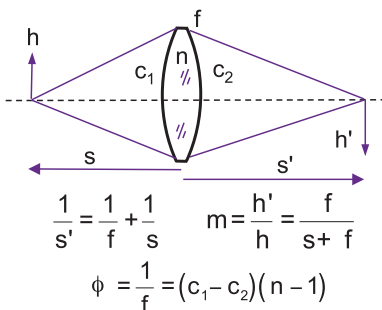
The powered surface of a **plano-convex singlet** made from glass ($n \sim 1.5$) has a convex radius ap-proximately equal to half its focal length. For the best performance, the sur-face should be oriented with the convex side facing the collimated light or laser beam.
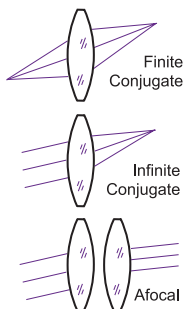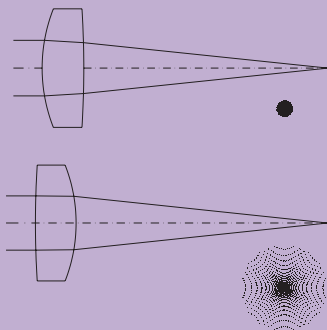
Objects and images form **conjugate** planes. A **finite conjugate** system has both a finite object and a finite image distance. An **infinite conjugate** system has either the object or the image located at infinity. In an **afocal** system, both the object and image are located at infinity. The performance of a lens designed for a specific conjugate usually degrades when used at other conjugates.

Finite Conjugate

Infinite Conjugate

Afocal

Aperture Stop and Field Stop
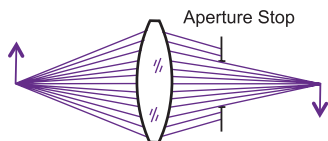
The **aperture stop** (often shortened to just **stop**) or **stop surface** is the *single* aperture in the lens system that limits the bundle of light that propagates through the system from the *axial* object point. In the absence of vignetting, the ray bundles from other object points also fill the aperture stop ("pivoting" about the center of the stop). The aperture stop is typically a mechanical surface (for good edge definition) and can be located anywhere in the optical system except at the object, the image, or an internal image. The size of the aperture stop sets the system $f/\#$ and NA. For objects at infinity, the aperture stop limits the input beam diameter.

A **ray fan** is a "slice" of rays that emerge from a single field point and pass through a line in the pupil. The **tangential** fan (Y-fan or **meridional** fan) passes through the pupil's *y* axis, and the **sagittal** fan (or X-fan) passes through the pupil's *x* axis.
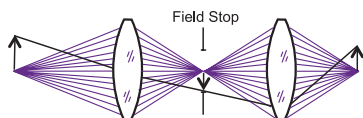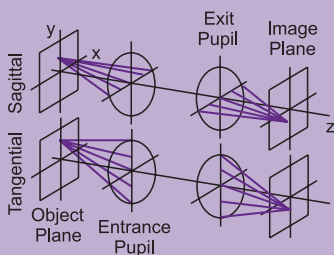
Field stops also limit the amount of light that can propagate through the system from the object plane to the image plane by setting the maximum object height from which a ray bundle can emerge and still pass through the lens. Field stops are typically the image (or the object) surface; however, if the system has an internal image plane, a field stop may be located there (usually for stray light control), and because this plane is conjugate with the image, reticles, crosshairs, or linear scales are placed at the internal field stop to be simultaneously imaged on the focal plane.

Entrance and Exit Pupils

Wavefront performance for imaging lenses is often measured at the exit pupil. The exit pupil and the image plane form a Fourier transform pair under Fresnel propagation. Optical design codes calculate **point spread functions (PSFs)** and other diffraction-based metrics using a **fast-Fourier transform (FFT)** of the exit pupil wavefront.

The physical stop surface is often buried in a lens system and difficult to access, leading to the definition of pupils. The images of the aperture stop formed by the lens elements before and after the stop surface are called the **entrance** and **exit pupils**, respectively. If the stop is at the first (last) surface, it is also the entrance (exit) pupil. The entrance (exit) pupil appears to limit the incoming (outgoing) bundle of rays in object (image) space. **Pupil matching** is a critical consideration when combining two independently designed optical systems. The entrance pupil of the second system should have the same position and diameter as the exit pupil of the first system to avoid losing light and reducing the FOV through vignetting. The entrance and exit pupils are images of each other forming conjugate planes, and, consequently, they can have **pupil aberrations**.
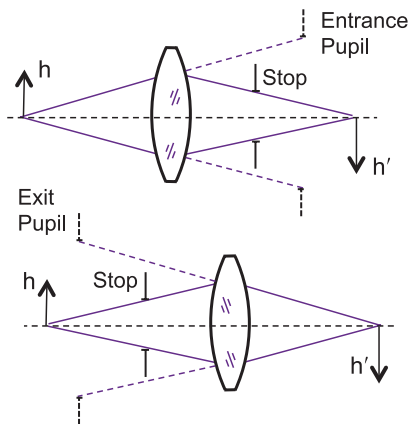
Similar to entrance and exit pupils, internal field stops can be imaged by the lens elements on either side of the field stop, forming **entrance** and **exit windows**. However, not all lenses have internal images and, therefore, entrance and exit windows.

Marginal and Chief Rays

Given the surface curvatures, material indices, and element thicknesses, the **marginal** and **chief rays** are the only two **reference rays** that need to be traced to calculate all first-order properties and third-order aberrations of a lens system.

The marginal ray starts at the base of the object (the on-axis, $H = 0$, field point) and passes through the edge of the aperture stop. The marginal ray locates all image positions, including any intermediate images; it also defines the edge of the ray bundle, and, therefore, the maximum sizes of the entrance and exit pupils.

The chief ray originates from the edge of the object (the full-field, $H = 1$, field point) and passes through the center of the aperture stop. The chief ray defines the maximum FOV of the lens system.

	Marginal	Chief
Ray Height	y, y_a	$\bar{y}, y_b$
Optical Axis	u, u_a	$\bar{u}, u_b$
Sfc. Normal	i, i_a	$\bar{\imath}, i_b$
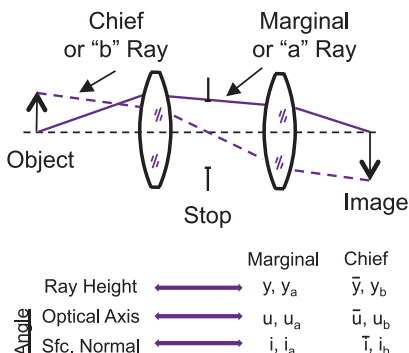
Pupils exist wherever the chief ray crosses the optical axis (the entrance and exit pupil are typically virtual; the chief ray needs to be extended "back" until it crosses the optical axis). The image size is determined by the height of the chief ray at the image plane.

The chief ray intercept on the focal plane may not define the center of the energy distribution for a given field point. In this case, the spot **centroid** or "center of mass" is a better metric for determining true field position at an image plane. Optical design programs allow a designer to reference various performance metrics (e.g., RMS spot diameter, distortion) from either the chief ray location or the centroid location.

Zernike Polynomials

Z_4	Z_5	Z_6	Z_7	Z_8	Z_9
(power)	(Astig. X)	(Astig. Y)	(Coma X)	(Coma Y)	(Spherical)
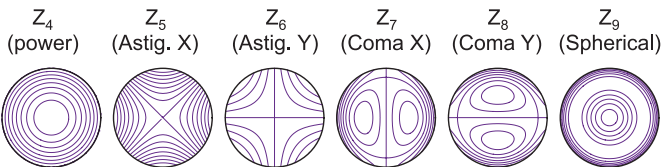
Zernike polynomials are a set of circularly symmetric orthogonal basis functions defined over a unit circle. They are 2D functions of both radial and azimuthal coordinates. In optical design, Zernikes are used to describe either surface irregularity or system wavefront (measured in the pupil). **Fringe Zernikes** are defined such that each polynomial has a value of unity at the edge of the unit circle; typically, only a limited 37-term subset is used, starting at Z_1. **Standard Zernikes** are numbered differently (starting with Z_0) and are used when many more coefficients are needed (e.g., simulations of turbulence or time-varying statistical phase surfaces). However, either type is capable of representing the same deformation. Standard Zernikes are normalized such that each polynomial term has an RMS value of 1 over the unit circle, allowing coefficients to add in quadrature to determine the overall RMS value. Both types of Zernikes can be renormalized so that they are orthogonal over an annular or otherwise noncircular region.

In tolerancing, Zernikes are useful for modeling complex **surface irregularity** (beyond cylinder and spherical) from polishing residuals or mounting effects.

When used to model wavefront error in the pupil, Zernikes appear identical to the traditional **Seidel aberrations**. However, Zernikes only represent the pupil wavefront at one field and are not true expansions of the **point characteristic** aberration function. Special Zernike expansions that are functions of both pupil and field have been developed for lithography applications.
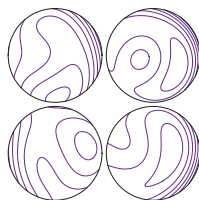
Conic Sections

A surface of rotation formed from a **conic section** is a special type of **asphere**. Conic surfaces are defined using the **sag equation** and parameterized by the conic constant κ. **Spheres**, **paraboloids**, **hyperboloids**, and **ellipsoids** are all surfaces formed from conic sections.
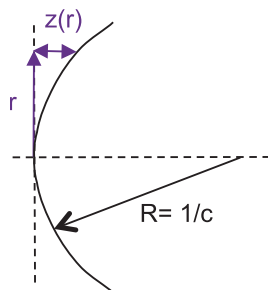
$$z(r) = \frac{cr^2}{1 + \sqrt{1 - (1+\kappa)(cr)^2}}$$

Hyperboloid ⟷ κ < −1
Prolate Ellipsoid ⟷ −1 < κ < 0
Paraboloid ⟷ κ = −1
Sphere ⟷ κ = 0
Oblate Ellipsoid ⟷ 0 > κ > 1

Conic surfaces have **stigmatic imaging** in which one point is perfectly focused to another point. In a sphere, the two focal points coincide. In a parabola, one point is at infinity. Hyperboloids have one real and one virtual focus, whereas ellipsoids have two real foci. The deterministic correspondence of foci makes testing conic surfaces easier than testing a general aspheric surface.

Prolate or "pointy" ellipsoids (football-shaped) are ellipses rotated about their major axis and have a negative κ. **Oblate** or "squashed" ellipsoids (pumpkin-shaped) are ellipses rotated about their minor axis with a positive κ. Oblate ellipsoids are more difficult to test.

Conic surfaces have closed-form analytic solutions for all third-order aberrations. Conic surfaces used off-axis are limited by coma and, when used at nonideal conjugates, suffer mainly from spherical aberration. Conic surfaces are very common in multiple-mirror reflective telescopes.
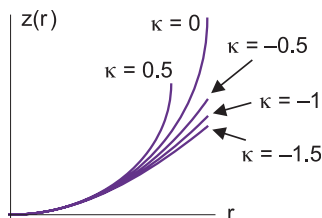
Diffraction Gratings

A **diffraction grating** splits in-
cident light into different angles
based on wavelength and **grating
spacing** d. For a given incident an-
gle, the exiting angle for order m
can be calculated with the **grat-
ing equation**. Because neither in-

$$m\lambda = d \left(\sin \theta_m - \sin \theta_i \right)$$

cident nor output angles can be greater than 90 deg, the
highest possible grating order is d/λ. Gratings can be
transmissive or reflective. A **Littrow grating** is a retrore-
flective grating with an output angle equal to the incident
angle for one particular diffraction order and wavelength.
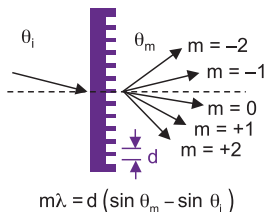
Tooling marks on diamond-turned surfaces can be
modeled as gratings for stray light analysis.

Linear gratings can be manufactured by etching or
scribing grooves on a glass/metal surface, photo-reducing
a regular pattern of lines on a glass plate, or by stamping
a master form into a substrate. **Curved gratings** can also
act as powered optical elements in complex optical systems
but are harder to fabricate and are usually reflective. A
holographic grating is formed by interfering two optical
beams in a photosensitive material.

Light in nondesired diffracted orders can cause significant
stray light. **Diffraction efficiency** is the ratio of power
in a given order to the total incident power. It is primarily
a function of grating order and grating "tooth" shape
(e.g., continuous or stepped). **Blazed gratings** have a
high diffraction efficiency in one order at the design or
blaze wavelength. Grating **dispersion** is the derivative
of output angle with respect to wavelength and is another
measure of grating efficiency. Increasing the order m will
increase dispersion but reduce the diffraction efficiency.

Free spectral range is the
total range of wavelengths in
a given order that do not
angularly overlap those of
another order.

$$\lambda_{blaze} = 2d \sin \alpha$$
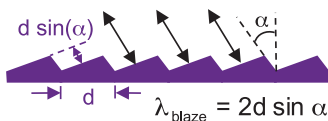
Optical Cements and Coatings

Achromatic doublets can be airspaced or **cemented** (where an **optical cement** is used to bond the achromat at the common internal surface). Airspaced doublets have significantly increased sensitivity to manufacturing tolerances, so cementing is the preferred option if a transparent cement exists for the wavelength range (e.g., cements that work well with low absorption in the UV and IR are difficult to find). Canada **balsam** (n = 1.53 to 1.55) and UV-cured synthetic adhesives are common in the visible and work over large temperature ranges. The cement layer must accommodate element CTE mismatches over the full temperature range that the lens is expected to see. A typical thickness for the cement layer is 10–20 μm and is usually not included in the optical design; however, very high-performing lenses may need to be optimized with the cement layer in place. Cemented surfaces do not need AR coatings unless the index difference between the glasses is large.

High-reflection (HR) coatings are needed on reflective optics to improve the base metal reflectivity (e.g., an **enhanced dielectric** coating can boost the reflectance of an aluminum mirror from 90% to over 97%). **Dichroics** are short- or long-wave-pass filters used for color balancing or color splitting. **Cold** and **hot mirrors** are special dichroics that transmit or reflect, respectively, thermal (long-wave) radiation to protect thermally sensitive optics. Cube beamsplitters require AR

coatings on the entrance and exit faces as well as an internal coating to divide the beam. **Polarizing beamsplitters** split unpolarized incident light into two orthogonally polarized beams.
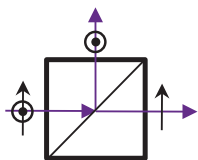
Reflective dielectric coatings generally have very low absorption, whereas metal coatings have more absorption. Some reflectors use **all-dielectric mirrors** to entirely avoid the absorption from the metal surfaces. Such mirrors may also provide spectral selectivity.

Detectors: Sampling

Modern lenses are used with periodically sampled digital staring arrays. The **Shannon sampling theorem** requires a minimum of two samples per cycle to perfectly reconstruct a sampled signal. The **Nyquist limit** defines the maximum frequency at which features can be sampled and corresponds to one full white–dark cycle spanning two pixels. Therefore, the highest spatial frequency that can be detected equals the reciprocal of twice the **pixel pitch** p. Real systems reliably resolve about 70% of the **Nyquist frequency** (roughly 2.8 pixels/cycle). Image content above
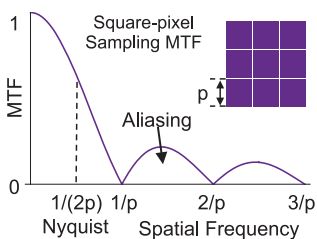
the **sampling frequency** ($1/p$) will **alias** and appear as lower-frequency content. An approximate sinc-function **sampling MTF** is often used for system modeling and assumes spatially averaged random-phase frequency content.

Not all detectors have rectangular geometry. Fiber optic faceplates have hexagonal spacing, leading to a sampling MTF with non-Cartesian symmetry.

For the best-performing optical system, the lens and the detector should be "matched" to each other. A common design trade is to optimize the ratio of the diffraction-limited optical blur spot to the sampling size. This

$$Q = \frac{\lambda(f/\#)}{\text{pixel pitch}}$$

ratio is referred to as the **Q parameter**, which is a comparison of the **Nyquist frequency** with the lens **cutoff frequency**. A Q value of 1 corresponds roughly to 2 pixels per Airy disc diameter. Q values range from 1–2 for most sensors; higher values of Q correspond to lens-limited (**oversampled**) imaging, and Q values under 1 correspond to pixel-limited (**undersampled**) imaging.
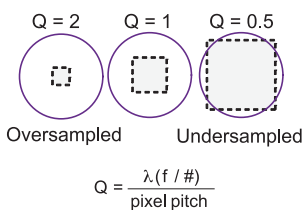
Detectors: Resolution

Limiting resolution (LR) is the spatial (or angular) frequency at which an MTF degrades to a desired contrast level. The lens is only one part of a complex optical system; besides the diffraction limit and lens aberrations, the detector is often the most significant contributor to the system MTF and limiting resolution. In CCDs for high-performance systems, a **charge-diffusion MTF** represents a blurring function caused by photoelectrons diffusing into neighboring pixels. Linear systems theory allows multiplication of cascaded component MTFs to obtain the final system MTF.

> Limiting resolution is often measured *subjectively* using a square-wave bar target in units of line-pairs/mm (lp/mm). LR *objectively* derived from an MTF curve has units of cy/mm. Although square-waves and sinusoids are related by an infinite Fourier series, cycles/mm and lp/mm should not be used interchangeably.

Both periodically sampled detectors (CCD/CMOS arrays) and film have MTF curves with finite resolution. Classic film LR (based on chemistry and grain size) can range from 80 cycles/mm for panchromatic film to over 2000 cy/mm for scientific-grade spectroscopic plates. LR for a digital array is typically defined by the **Nyquist frequency**, equal to the reciprocal of twice the pixel pitch. The Nyquist frequency for a 3-μm pixel is 167 cycles/mm, comparable to high-performance photographic film.

Limiting resolution is insufficient to determine overall *system* image quality because two very different system MTFs can have the same LR. A complete MTF curve is a much better performance predictor.

Spatial frequency (cycles/mm)
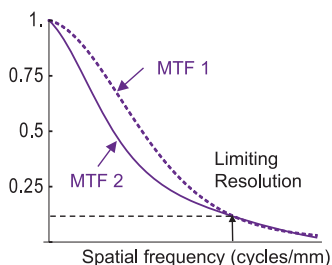
The Lagrange Invariant and Étendue

In an optical system, the **Lagrange invariant** (also called the **optical invariant**) is a constant that can be calculated at any surface within the lens. At the image surface, it equals the product of the image space index, the image height, and the marginal ray (convergence) angle. At the object surface, it equals the product of the object space index, the object height, and the marginal ray (divergence) angle. At other surfaces, it is given by a combination of marginal and chief ray angles and heights:

$$H = n\bar{u}y - nu\bar{y} \qquad n^2 A\Omega = \pi^2 H^2.$$

For a fixed image size and $f/\#$, the object size cannot be reduced (increased) without illuminating the object over a larger (smaller) NA.

Étendue (or throughput) is the 3D optical invariant and equals the product of the image (object) area times the solid angle of collection (illumination). Étendue is proportional to the square of the Lagrange invariant; it is
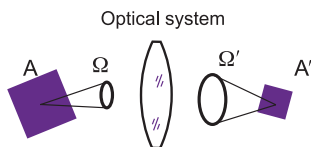
Optical system

Etendue = $\iint_{\substack{\text{Any} \\ \text{surface}}} dA d\Omega$

a geometric measure of the total **brightness** of an optical system. The conservation of radiance states that the étendue of an optical bundle can only remain constant (or increase) through an optical system—it cannot decrease and still obey energy conservation. Étendue is purely geometric and has no inherent relation to imaging performance, representing only the boundaries of the light bundle and not the actual distribution. An aberrated blur spot will *increase* étendue because the beam area is larger for the same solid angle.

For a flat surface with a uniform solid angle, the étendue equals $\pi A \sin^2(\theta)$, where A is the area of the surface, and θ is the half angle of the marginal ray.

Illumination Design

Many imaging systems also require optimized **illuminators** (e.g., projectors, microscopes, and machine vision systems). The primary goal of **illumination design** is to create a desired spatial and/or angular distribution of energy at a given surface. Illumination systems are also found in automobile headlights, spotlights, commercial lighting, and displays. **Collection** or **concentrator** systems maximize radiation transfer between a bright source and a target surface, such as a simple parabolic or elliptical concentrator.
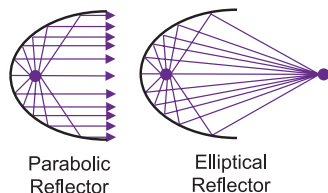
Parabolic Reflector

Elliptical Reflector

Creating a uniform distribution from a complex source requires **homogenization**, which can be done by **tailoring** a known source distribution or by using **superposition** of many source images. Automotive headlights and streetlights use wedged parabolas as tailored homogenizers. **Light pipes** and **fly's eye arrays** are examples of superposition homogenizers.

Critical (or **Abbe**) **illumination** images the source directly onto the image, increasing brightness but potentially adding unwanted image structure (e.g., the lamp filament). In contrast, **Köhler illumination** images the source into the imaging system pupil for a more uniform illumination but requires more lenses.

In a microscope, **bright field illumination** fills the imaging pupil either through the objective (**epi-illumination**) or from underneath the sample (**trans-illumination**). **Dark field illumination** puts no light in the imaging pupil—the imaging system collects scattered light, highlighting edges. Lithography illuminators fill the pupil with complex ring or dipole illumination. The illumination **coherence** can play a critical role in high-performance systems by shifting the performance limits (partially coherent light has a slightly different "diffraction-limited" spot size than fully coherent light).

Equation Summary

Numerical aperture and *f*/#:

$$NA = n' \sin u' \qquad f/\# = \frac{f}{EPD}$$

Rayleigh criterion:

$$\Delta X = \frac{0.61\lambda}{NA}$$

Image height as a function of field angle:

$$h' = f \tan \theta$$

Transverse ray error:

$$\varepsilon'_y = \frac{1}{n'u'_a} \frac{\partial W}{\partial \rho_y} \qquad \varepsilon'_x = \frac{1}{n'u'_a} \frac{\partial W}{\partial \rho_x}$$

Strehl ratio:

$$Strehl \cong \left(1 - 2\pi^2 \omega^2\right) \qquad \omega = RMS_{OPD}$$

Wavefront aberration polynomial:

$$W_{IJK} \Rightarrow H^I \rho^J \cos^K \theta$$

$$W(H, \rho, \theta) = W_{020}\rho^2 + W_{111}H\rho\cos\theta + W_{040}\rho^4 + W_{131}H\rho^3\cos\theta$$
$$+ W_{222}H^2\rho^2\cos^2\theta + W_{220}H^2\rho^2 + W_{311}H^3\rho\cos\theta + O(6)$$

Contrast:

$$Contrast = \frac{I_{\max} - I_{\min}}{I_{\max} + I_{\min}}$$

Equation Summary

Focal lengths of any two thin lens system:

$$f_a = \frac{d\,f}{f - BFL} \qquad f_b = \frac{d\,BFL}{f - BFL - d}$$

Zero-Petzval solution for two thin lenses:

$$f_a = -f_b = f - BFL \qquad d = \frac{(f - BFL)^2}{f}$$

Two-mirror solution:

$$c_1 = \frac{BFL - f}{2\,d\,f} \qquad c_2 = \frac{BFL + d - f}{2\,d\,BFL}$$

Schwarzchild solution:

$$d = 2f \qquad c_1 = \left(\sqrt{5} - 1\right)f \qquad c_2 = \left(\sqrt{5} + 1\right)f$$

Aplanatic condition:

$$i = -u'$$

Bending and shape factors:

$$\beta = \frac{c_1 + c_2}{c_1 - c_2} \qquad C = \frac{u_a + u_a'}{u_a - u_a'}$$

Thin lens bending for minimum spherical:

$$\frac{c_2}{c_1} = \frac{2n^2 - n - 4}{n(2n + 1)}$$

Equation Summary

Thin lens bending for minimum coma:

$$\frac{c_2}{c_1} = \frac{(n^2 - n - 1)}{n^2}$$

Achromatic doublet:

$$\Phi = \phi_1 + \phi_2 \qquad \phi_1 = \Phi \frac{V_1}{V_1 - V_2} \qquad \phi_2 = -\Phi \frac{V_2}{V_1 - V_2}$$

Petzval sum:

$$\sum_j \frac{\phi_j}{n_j}$$

Thick lens power:

$$\phi = \phi_1 + \phi_2 - \frac{t}{n}(\phi_1 \phi_2)$$

Minimum clear aperture for no vignetting:

$$\mathrm{CA}_{\min} = |y_a| + |y_b|$$

Aspheric sag equation:

$$sag = z(r) = \frac{c\,r^2}{1 + \sqrt{1 - (\kappa + 1)(c\,r)^2}} + d\,r^4 + e\,r^6 + f\,r^8 + g\,r^{10} + \ldots$$

Gradient index profiles:

$$n(r) = N_{00} + N_{10}\,r^2 + N_{20}\,r^4 + \ldots$$
$$n(z) = N_{00} + N_{01}\,z + N_{02}\,z^2 + \ldots$$

Equation Summary

Merit function:

$$\phi = \sum_{i=1}^{m} w_i^2 (c_i - t_i)^2$$

Athermalization condition:

$$\frac{df_{Lens}}{dT} = CTE_1 d_1 \quad CTE_2 d_2$$

Bireflectance scattering distribution function:

$$BSDF(\theta_i, \phi_i; \theta_o, \phi_o) = \left(\frac{L}{E}\right) sr^{-1}$$

Sellmeier dispersion:

$$n(\lambda) = \sqrt{1 + \frac{c_1 \lambda^2}{\lambda^2 - c_4} + \frac{c_2 \lambda^2}{\lambda^2 - c_5} + \frac{c_3 \lambda^2}{\lambda^2 - c_6}}$$

Schott dispersion:

$$n(\lambda) = \sqrt{c_1 + c_2 \lambda^2 + \frac{c_3}{\lambda^2} + \frac{c_4}{\lambda^4} + \frac{c_5}{\lambda^6} + \frac{c_6}{\lambda^8}}$$

Snell's law:

$$n \sin\theta = n' \sin\theta'$$

Paraxial ray tracing:

$$n' u' = nu - y\phi$$
$$y' = y + n' u' \left(\frac{d}{n'}\right)$$
$$\phi = c(n' - n)$$

Field Guide to Lens Design

Equation Summary

Lens maker's equation and linear magnification:

$$\frac{1}{s'} = \frac{1}{f} + \frac{1}{s} \qquad m = \frac{h'}{h} = \frac{f}{s+f}$$

Thin lens power:

$$\Phi = \frac{1}{f} = (c_1 - c_2)(n-1)$$

Diffraction gratings:

$$m\lambda = d\left[\sin(\theta_m) - \sin(\theta_i)\right] \qquad \lambda_{blaze} = 2d \sin\alpha$$

Sampling ratio:

$$Q = \frac{\lambda(f/\#)}{pixel\ pitch}$$

Lagrange invariant and étendue:

$$H = n\,\bar{u}\,y - n\,u\,\bar{y} \qquad n^2 A\Omega = \pi^2 H^2$$

$$Etendue = \iint\limits_{surface} dA\,a\Omega$$

Bibliography

G. Boreman, *Modulation Transfer Function in Optical and Electro-Optical Systems*, SPIE Press, Bellingham, WA (2001) [doi:10.1117/3.419857].

B. Braunecker, R. Hentschel, and H. J. Tiziani, Eds., *Advanced Optics Using Aspherical Surfaces*, SPIE Press, Bellingham, WA (2008) [doi:10.1117/3.741689].

R. Fischer, B. Tadic-Galeb, and P. Yoder, *Optical System Design*, 2nd ed., McGraw-Hill, New York (2008).

G. W. Forbes, Shape specification for axially symmetric optical surfaces, *Opt. Express* **15**(8), 5218–5226 (2007).

J. M. Geary, *Introduction to Lens Design*, Willmann-Bell, Richmond, VA (2002).

J. E. Grievenkamp, *Field Guide to Geometrical Optics*, SPIE Press, Bellingham, WA (2003) [doi:10.1117/3.547461].

H. Gross, Ed., *Handbook of Optical Systems*, Vol. 1–4, John Wiley and Sons, New York (2005).

H. Karow, *Fabrication Methods for Precision Optics*, John Wiley and Sons, New York (1993).

M. J. Kidger, *Fundamental Optical Design*, SPIE Press, Bellingham, WA (2002) [doi:10.1117/3.397107].

M. J. Kidger, *Intermediate Optical Design*, SPIE Press, Bellingham, WA (2004) [doi:10.1117/3.540692].

R. Kingslake, *Optical System Design*, Academic Press, New York (1983).

R. Kingslake and B. Johnson, *Lens Design Fundamentals*, 2nd ed., Academic Press, New York (2010).

M. Laiken, *Lens Design*, Marcel Dekker, Inc., New York (1995).

D. Malacara, *Optical Shop Testing*, 3rd ed., John Wiley and Sons, New York (2007).

Bibliography

D. O'Shea, *Elements of Modern Optical Design*, John Wiley and Sons, New York (1985).

R. R. Shannon, *The Art and Science of Optical Design*, Cambridge University Press, Cambridge, UK (1997).

G. H. Smith, *Camera Lenses: From Box Camera to Digital*, SPIE Press, Bellingham, WA (2006).

W. Smith, *Modern Lens Design*, 2nd ed., McGraw-Hill, New York (2005).

W. Smith, *Modern Optical Engineering*, 3rd ed., McGraw-Hill, New York (1996).

B. H. Walker, *Optical Design for Visual Systems*, SPIE Press, Bellingham, WA (2000) [doi:10.1117/3.391324].

W. T. Welford, *Aberrations of Optical Systems*, Taylor and Francis, New York (1986).

C. S. Williams and O. Becklund, *Introduction to the Optical Transfer Function*, SPIE Press, Bellingham, WA (2002).

W. L. Wolfe, *Introduction to Infrared System Design*, SPIE Press, Bellingham, WA (2000) [doi:10.1117/3.226006].

P. Yoder, *Opto-Mechanical Systems Design*, 3rd ed., Taylor and Francis, New York (2006).

Index

Index

Index

Index

Index

Index

Index

Index

Index

Index

Index

Index

Julie Bentley is an Associate Professor at The Institute of Optics, University of Rochester and has been teaching courses in geometrical optics, optical design, and product design for more than 15 years. She received her B.S., M.S., and Ph.D. in Optics from the University of Rochester. After graduating, she spent two years at Hughes Aircraft Co. in California and then twelve years at Corning Tropel in New York. She has experience designing a wide variety of optical systems for both commercial and military markets, from the UV to the IR, in both refractive and reflective configurations. She now owns Bentley Optical Design, an optical design consulting company. Dr. Bentley is a Member of OSA and an SPIE Fellow.

Craig Olson is a Principal Engineer at L-3 Communications. He received a B.S. in Electrical Engineering from the Georgia Institute of Technology and a Master's and Ph.D. in Optics from the Institute of Optics at the University of Rochester. He previously worked in the telecommunications component sector and the commercial product sectors at JDS Uniphase. In addition to developing sensor system models, he currently designs, builds, and tests active and passive multispectral sensor payloads for airborne and ground-based imaging systems—occasionally finding time to design a lens. His current research interests involve the physics of image formation throughout the complete imaging chain. Dr. Olson is a Member of OSA and IEEE, and a Senior Member of SPIE.